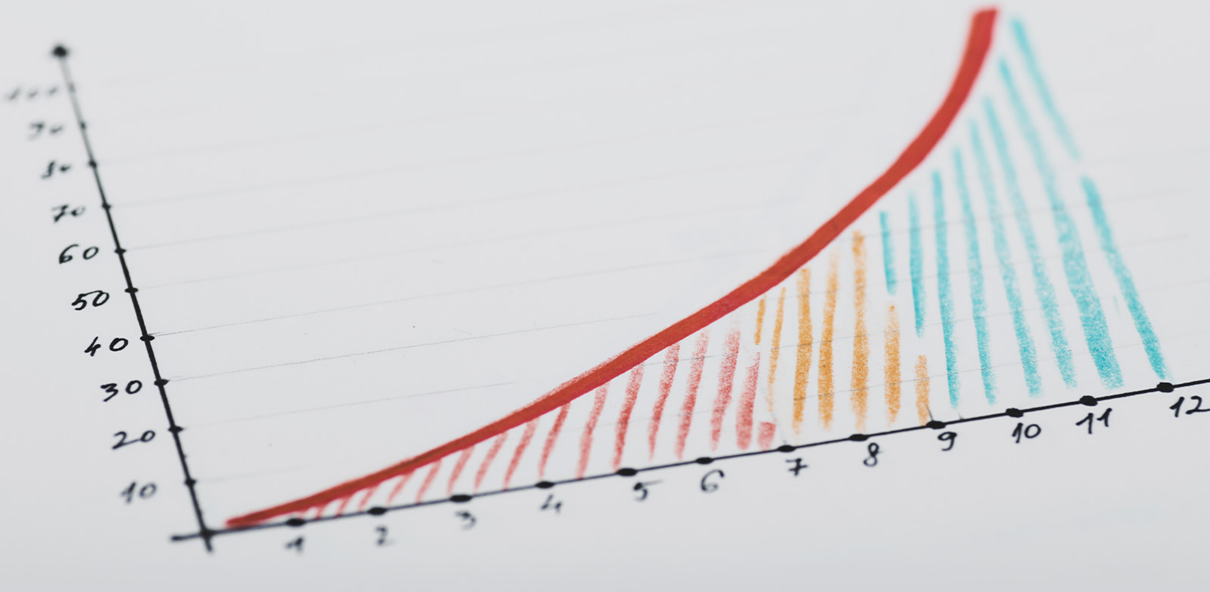
Mirian Alexandra Valeriano Meneses
Santiago Andrés Otero-Potosi

GUIA GENERAL DE ESTUDIO DE LA ASIGNATURA DE

ESTADÍSTICA APLICADA PARA EDUCACIÓN INICIAL

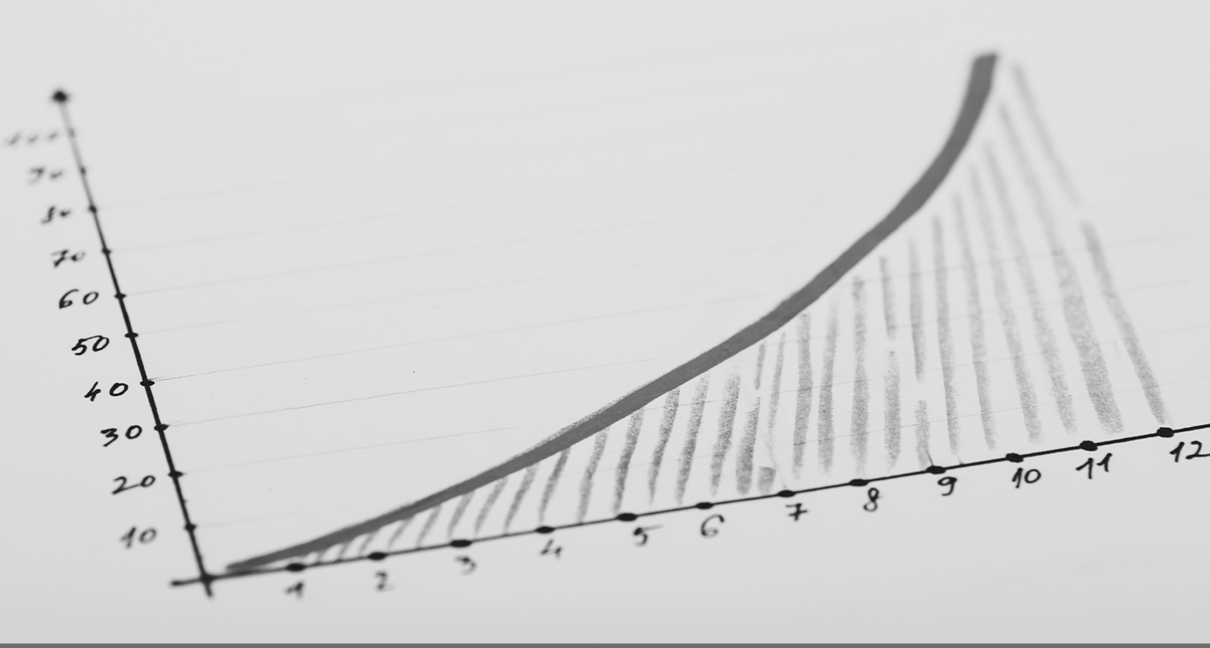


Atena Editora
Año 2024

Mirian Alexandra Valeriano Meneses
Santiago Andrés Otero-Potosi

GUIA GENERAL DE ESTUDIO DE LA ASIGNATURA DE

ESTADÍSTICA APLICADA PARA EDUCACIÓN INICIAL



Atena Editora
Año 2024

Editora jefe
Profª Drª Antonella Carvalho de Oliveira
Editora ejecutiva
Natalia Oliveira
Asistente editorial
Flávia Roberta Barão
Bibliotecario
Janaina Ramos
Proyecto gráfico
Ellen Andressa Kubisty
Luiza Alves Batista
Nataly Evilin Gayde
Thamires Camili Gayde
Imágenes de portada
iStock
Edición de arte
Luiza Alves Batista

Todos los manuscritos fueron previamente sometidos a evaluación ciega por pares, miembros del Consejo Editorial de esta editorial, habiendo sido aprobados para su publicación con base en criterios de neutralidad e imparcialidad académica.

Atena Editora se compromete a garantizar la integridad editorial en todas las etapas del proceso de publicación, evitando plagios, datos o entonces, resultados fraudulentos y evitando que los intereses económicos comprometan los estándares éticos de la publicación. Las situaciones de sospecha de mala conducta científica se investigarán con el más alto nivel de rigor académico y ético.

Guia general de estudio de la asignatura de estadística aplicada para educación inicial

Diagramación: Ellen Andressa Kubisty
Correción: Maiara Ferreira
Indexación: Amanda Kelly da Costa Veiga
Revisión: Los autores
Autores: Mirian Alexandra Valeriano Meneses
Santiago Andrés Otero-Potosi

Datos de catalogación en publicación internacional (CIP)

M543 Meneses, Mirian Alexandra Valeriano
Guia general de estudio de la asignatura de estadística aplicada para educación inicial / Mirian Alexandra Valeriano Meneses, Santiago Andrés Otero-Potosi. – Ponta Grossa - PR: Atena, 2024.

Formato: PDF
Requisitos del sistema: Adobe Acrobat Reader
Modo de acceso: World Wide Web
Incluye bibliografía
ISBN 978-65-258-2850-3
DOI: https://doi.org/10.22533/at.ed.503243008

1. Estadística. 2. Educación inicial. 3. Probabilidad. 4. Desarrollo cognitivo. 5. Aprendizaje interactivo. I. Meneses, Mirian Alexandra Valeriano. II. Otero-Potosi, Santiago Andrés. III. Título.

CDD 310

Preparado por Bibliotecario Janaina Ramos – CRB-8/9166

Atena Editora
Ponta Grossa – Paraná – Brasil
Telefone: +55 (42) 3323-5493
www.atenaeditora.com.br
contato@atenaeditora.com.br

DECLARACIÓN DE LOS AUTORES

Los autores de este trabajo: 1. Certifican que no tienen ningún interés comercial que constituya un conflicto de interés en relación con el artículo científico publicado; 2. Declaran haber participado activamente en la construcción de los respectivos manuscritos, preferentemente en: a) Concepción del estudio, y/o adquisición de datos, y/o análisis e interpretación de datos; b) Elaboración del artículo o revisión para que el material sea intelectualmente relevante; c) Aprobación final del manuscrito para envío; 3. Acrediten que los artículos científicos publicados están completamente libres de datos y/o resultados fraudulentos; 4. Confirmar la cita y la referencia que sean correctas de todos los datos e interpretaciones de datos de otras investigaciones; 5. Reconocen haber informado todas las fuentes de financiamiento recibidas para la realización de la investigación; 6. Autorizar la publicación de la obra, que incluye las fichas del catálogo, ISBN (Número de serie estándar internacional), D.O.I. (Identificador de Objeto Digital) y demás índices, diseño visual y creación de portada, maquetación interior, así como su lanzamiento y difusión según criterio de Atena Editora.

DECLARACIÓN DEL EDITOR

La Estadística en el entorno está presente en la mayoría de las actividades, desde la perspectiva de la Docencia es imprescindible que, se domine el tratamiento de datos, por ejemplo, los resultados obtenidos en una evaluación y con la interpretación de los resultados se puedan tomar decisiones de fortalecimiento académico.

La educación inicial es crucial en la formación de los futuros ciudadanos, sentando las bases para el desarrollo de habilidades cognitivas y socioemocionales esenciales para el éxito en la vida. Es en este nivel educativo donde la estadística puede comenzar a desempeñar un rol fundamental, introduciendo a los niños en el mundo de los datos y el análisis de información de manera lúdica y significativa.

La presente obra pretende fortalecer el aprendizaje de Estadística desde el contexto educativo respecto a la Educación en la primera infancia, donde los futuros docentes cuenten con una guía para su aprendizaje. La guía cuenta con ejemplos ilustrativos para la mejor comprensión, los diferentes ejercicios están contextualizados y pretenden ser un referente para la construcción de sus propios ejemplos.

Los Autores

SUMARIO

SUMARIO

UNIDAD I

ESTADÍSTICA DESCRIPTIVA

¿QUÉ ES LA ESTADÍSTICA?

La Estadística es fundamental en investigación y análisis de datos, su finalidad es recopilar, organizar, interpretar y presentar datos numéricos para describir fenómenos, realizar inferencias y tomar decisiones informadas, esta ciencia se aplica en muchos campos, desde la economía hasta la medicina, pasando por la ingeniería y las ciencias sociales, mediante métodos estadísticos, los investigadores pueden obtener conclusiones significativas a partir de conjuntos de datos complejos, identificando patrones, tendencias y relaciones que ayudan a comprender mejor el mundo que nos rodea (Hernández et al., 2023).

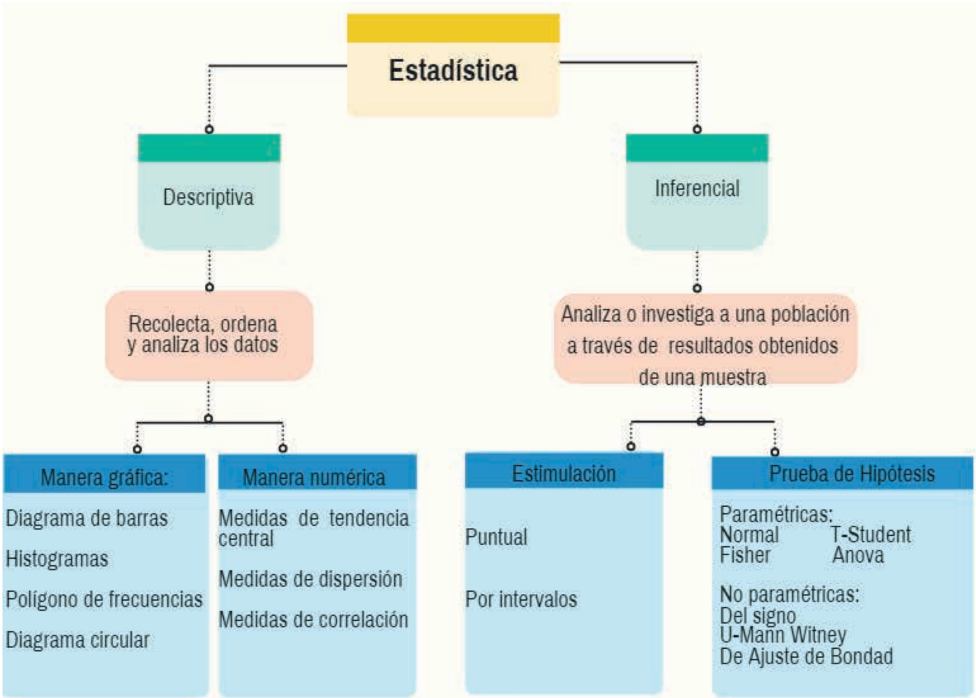
Una parte esencial es la recolección sistemática y precisa de datos mediante encuestas, experimentos y observaciones para recopilar información relevante sobre variables de interés. Posteriormente, los datos se organizan en tablas, gráficos o distribuciones que permiten una visualización clara y facilitan su análisis, siendo la calidad de los resultados que depende de la eficacia y la precisión de los parámetros recolectados, por lo que es crucial emplear técnicas rigurosas en esta etapa inicial (Casas et al., 2003).

Otro aspecto fundamental de la estadística es la inferencia, que implica hacer generalizaciones o predicciones sobre una población basándose en una muestra representativa de esa localidad, esta técnica se utiliza para sacar conclusiones más allá de los datos específicos observados y para estimar parámetros desconocidos, la inferencia estadística se basa en principios probabilísticos y requiere el uso adecuado de métodos como intervalos de confianza, pruebas de hipótesis y análisis de regresión para obtener resultados significativos y válidos (Yanqui et al., 2024).

Además de su papel en la investigación académica y científica, la estadística desempeña un papel crucial en la toma de decisiones en el ámbito empresarial y gubernamental, las empresas utilizan análisis estadísticos para evaluar el rendimiento, prever tendencias de mercado y optimizar procesos, del mismo modo, los gobiernos utilizan datos estadísticos para formular políticas públicas, asignar recursos y monitorear indicadores clave de desarrollo económico y social, a continuación se presenta un organizador gráfico que resume de manera general la estadística:

Figura 1

Modelo de una matriz de responsabilidad RACI



Nota: Esta figura representa un ejemplo de una matriz de responsabilidad RACI, misma que fue elaborada por los autores, 2024

CONCEPTOS CLAVE

Población

En el contexto de la estadística, es fundamental entender la distinción entre poblaciones finitas e infinitas, ya que esta diferencia afecta la forma en que se abordan los problemas de muestreo y la inferencia estadística.

Población finita: un conjunto definido y limitado de elementos o individuos que comparten una característica común y que se pueden contar o enumerar de manera precisa, en este tipo, cada miembro se conoce y puede identificar; por ejemplo, la población de estudiantes inscritos en una escuela durante un año académico específico o la cantidad de empleados en una empresa constituyen ejemplos de poblaciones finitas, la característica de este tipo de población es que su tamaño total es conocido y finito, lo que permite analizar todos los elementos si es necesario, como en el caso de un censo (Castañeda et al., 2010).

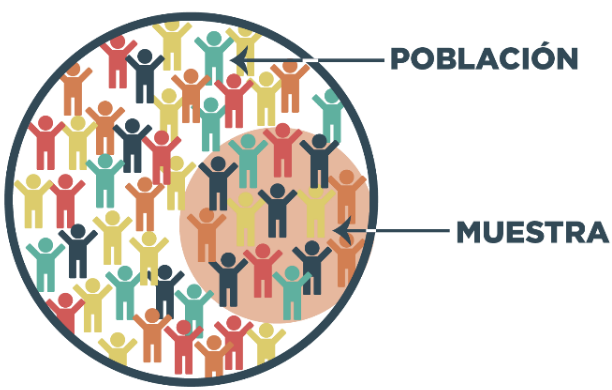
Población infinita: Una población infinita no tiene límite definido en términos de tamaño o extensión y no puede enumerarse o medirse en su totalidad, este tipo de población es conceptual y teórica, y ejemplos comunes incluyen la totalidad de los números naturales, la población mundial o la cantidad teórica de lanzamientos de una moneda, en el contexto estadístico, trabajar con este tipo de poblaciones implica que solo podemos acceder a una muestra limitada de elementos para realizar inferencias sobre la población en su conjunto; las técnicas de muestreo probabilístico que se utilizan para seleccionar muestras representativas que permitan generalizar conclusiones sobre la población (Guevara et al., 2020).

Muestra

Para ilustrar cómo se aborda el estudio de una población infinita en estadística, consideremos estimar la proporción de personas que prefieren un cierto producto en una ciudad grande, ya que la población de la ciudad es casi incontable, es imposible encuestar a cada individuo, en cambio, se selecciona una muestra representativa con técnicas de muestreo probabilístico, como el muestreo aleatorio simple (Alban et al., 2020).

Figura 2

Representación gráfica de la población



Nota: Esta figura muestra representativamente como se considera una muestra a partir de una población

CÁLCULO MUESTRAL

Para calcular el tamaño de la muestra y esta sea fiable, es necesario emplear las siguientes fórmulas:

Para una muestra con Población Finita

$$n = \frac{N * Z^2 * p * q}{e^2 * (N - 1) + Z^2 * p * q}$$

Para una muestra con Población Infinita

$$n = \frac{Z^2 * p * q}{e^2}$$

Donde:

n = Tamaño de la muestra que se busca

N = Tamaño de la Población

Z = Parámetro de confianza obtenido de los niveles de confianza**,** el valor recomendable cuando se desconoce es del 95% entonces Z= 1,96

e = Error de estimación aceptado, cuando no se conoce su valor queda al criterio del investigador, el valor generalmente varía entre 1% y 9%, lo recomendable es trabajar con el valor al 3% (0,03) y 5% (0,05).

p = Probabilidad de éxito que ocurra el evento, cuando no se conoce se trabaja con el 50% cuyo valor es (0,5).

q = Probabilidad de que No ocurra el evento, cuando no se conoce se trabaja con el 50% cuyo valor es (0,5).

DATO ESTADÍSTICO

Un dato estadístico es un valor numérico que describe alguna característica o propiedad de interés en un conjunto de datos, estos valores son fundamentales en el campo de la estadística, ya que proporcionan información cuantitativa que puede ser analizada y utilizada para tomar decisiones informadas, los datos estadísticos pueden ser recopilados a través de diferentes métodos como encuestas, experimentos o registros administrativos, y su análisis permite identificar patrones, tendencias y relaciones entre variables, ejemplos comunes de datos estadísticos incluyen la edad promedio de una población, la tasa de desempleo en una región específica, o la cantidad de ventas de un producto durante un periodo de tiempo determinado (Pérez et al., 2012).

VARIABLE ESTADÍSTICA

Una variable estadística es una característica o atributo que puede tomar diferentes valores en un estudio o experimento, estas pueden ser de dos tipos principales: variables cuantitativas, que representan cantidades numéricas y pueden ser medidas con precisión (como la edad de una persona o el ingreso anual), y variables cualitativas, que representan cualidades o categorías que no pueden ser medidas numéricamente (como el género, la profesión o el estado civil).

Las variables estadísticas son fundamentales porque permiten describir y analizar fenómenos de interés mediante la recolección y análisis de datos, dependiendo del tipo de variable y del objetivo del estudio, se utilizan diferentes métodos estadísticos para su análisis, como el cálculo de medidas de tendencia central (como la media o la mediana) o la realización de pruebas de hipótesis para determinar relaciones entre variables (Franco & Solórzano, 2020).

Figura 3

Clasificación de variables *estadísticas*



Nota: Esta figura representa un ejemplo de una matriz de responsabilidad RACI, misma que fue elaborada por los autores, 2024

DISTRIBUCIÓN DE FRECUENCIAS

La distribución de frecuencias es una técnica estadística utilizada para ordenar y resumir datos numéricos o categóricos en tablas que muestran la frecuencia con la que ocurren diferentes valores o categorías, contando el número de veces que ocurre cada valor en un conjunto de datos y organiza esta información en una tabla con categorías o rangos de valores junto con sus respectivas frecuencias (Fernández & Vela, 2021).

La distribución de frecuencias es útil para identificar patrones, tendencias y características destacadas de los datos, lo que facilita su análisis y comprensión, además, es posible calcular medidas descriptivas como la moda (el valor más frecuente), la mediana (el valor central) y la media (el promedio aritmético), que proporcionan información adicional sobre la distribución de los datos.

TABLAS DE FRECUENCIAS

Una tabla de frecuencias es una representación organizada y estructurada de datos que muestra la distribución de frecuencias de diferentes valores o categorías dentro de un conjunto de datos, esta tabla consta de columnas que representan los valores o intervalos de valores posibles y una columna adicional que muestra la frecuencia con la que cada valor o intervalo ocurre en los datos (Rumsey, 2002).

En cuanto a datos numéricos, los valores se agrupan en intervalos o clases para facilitar la interpretación y el análisis, la tabla de frecuencias da una visión clara y concisa de la distribución de los datos, identificando rápidamente los valores más comunes, los valores atípicos y las tendencias generales, además, a partir de una tabla, se pueden calcular medidas estadísticas importantes, como la moda, la mediana y la media.

Los tipos de frecuencia pueden ser					
Frecuencia Absoluta (f o fi)	**Frecuencia Relativa (fr)**	**Frecuencia absoluta acumulada (Fa)**	**Frecuencia Porcentual (f %)**	**Frecuencia relativa acumulada (Fr)**	**Frecuencia porcentual acumulada (Fr %)**
Es el número de veces que aparece o se repite el valor de la variable. La suma de las frecuencias absolutas es el número de datos N.	Indica la proporción con que se repite un valor. Se divide la frecuencia absoluta entre el número total de datos N. La suma de las fr es siempre 1. $fr = \frac{fi}{N}$	Es la suma acumulada de la frecuencia absoluta primera con la segunda, este valor con la tercer y así sucesivamente.	También llamada frecuencia relativa porcentual. Indica el valor de la **fr** en porcentaje, se obtiene multiplicando por 100% La suma de las frecuencias porcentuales es siempre 100%.	Es la suma de la frecuencia relativa primera con la segunda, este valor con la tercera, y así sucesivamente.	Se obtiene multiplicando la frecuencia relativa acumulada por 100%.

Ejemplo con datos No agrupados

Ordene los datos y construya la tabla de distribución de frecuencias para los siguientes datos:

Las notas de las evaluaciones parciales de 30 estudiantes en la asignatura de Estadística Aplicada son las siguientes:

8	7	5	10	9	9	8	7	10	6
10	10	9	5	7	8	8	7	9	9
7	7	9	9	8	8	10	5	6	5

Proceso de solución

- Como primer paso se debe ordenar los datos de la variable **calificación** en orden ascendente de menor a mayor
- Segundo, construir la tabla de frecuencias
- Posterior, contar las frecuencias de cada calificación
- Por último, realizar los cálculos respectivos

Calificación	Fi	Fr	Fa	f%	Fr	Fr%
5	4	4/30 = 0,133	4	0,133 *100% = 13,3 %	0,133	0,133*100% = 13,3%
6	2	2/30 = 0,066	4+2 = 6	0,066 *100% = 6,6 %	0,133 + 0,066= 0,199	0,199*100% = 19,9%
7	6	6/30 = 0,2	6+6 = 12	0,2 *100% = 20%	0,199 + 0,2 = 0,399	0,399*100% = 39,9%
8	6	6/30 = 0,2	12+6 = 18	0,2 *100% = 20%	0,399 + 0,2 = 0,599	0,599*100% = 59,9%
9	7	7/30 = 0,233	18+7 =25	0,233 *100% = 23,3%	0,599 + 0,233 = 0,832	0,832*100% = 83,2%
10	5	5/30 = 0,166	25+5 = 30	0,166 *100% = 16,6%	0,832 + 0,166 = 1	1*100% = 100%
Total	30	1		100%		

DISTRIBUCIÓN DE FRECUENCIAS PARA DATOS AGRUPADOS

Se usará esta tabla de frecuencias cuando la variable tiene una gran cantidad de datos o es una variable continua. Las agrupaciones de los datos se hacen en **intervalos** de igual amplitud a los cuáles se denominan **clases.**

Un **intervalo** es un conjunto de números contenidos entre dos extremos que se denominan límite inferior (valor extremo izquierdo) y límite superior (valor extremo derecho).

Pasos para construir una tabla de frecuencias para datos agrupados:

Cálculo del Rango R: es la diferencia entre el (valor mayor) con el (menor valor) del conjunto de datos.

$$R = X_{max} - X_{min}$$

Cálculo del número de intervalos K o ***I:*** es el número de subgrupos que se conformarán dentro del grupo de datos.

Para el cálculo del número de intervalos se sugiere emplear la fórmula de Sturges:

$$k = 1 + 3{,}322 * \log N$$

Donde:

N= es el número de datos

Por otro lado, también es recomendable emplear la raíz cuadrade de N, donde $k = \sqrt{N}$

Consideraciones

- Si el valor de k es 7,3 se redondea a 7 intervalos
- Si el valor de k es 6,64 se redondea a 7 intervalos

Cálculo de la amplitud A del intervalo: es el número de datos que contendrá cada intervalo de clase.

Se obtiene dividiendo el **rango (R)** para el **número de intervalos (k):**

$$A = \frac{R}{k}$$

Límites de clase: cada **clase** es un **intervalo** que va desde el **límite inferior (*Li*)** hasta el **límite superior (*Ls*).**

Cada intervalo: empieza cerrado [y finaliza abierto), se forma empezando por el (menor valor) de los datos y sumando la cantidad que indique la amplitud y se determina el límite superior.

Cálculo de la marca de clase X: es el **punto medio** de cada intervalo, se obtiene sumando el límite superior con el inferior y dividiendo entre 2.

$$X_m = \frac{Ls + Li}{2}$$

Pasos para construir una Tabla de frecuencias para datos agrupados

1. Determinar el rango (R)
2. Hallar el número de intervalos
3. Determinar la amplitud o ancho de clase
4. Hallar el límite inferior y superior de cada clase y sus marcas de clase
5. Ordenar los valores encontrados en las columnas de la tabla de distribución de frecuencias

Modelo de tabla

Intervalo de clase	Marcas de clase (x)	Frecuencia absoluta (f o fi)	Frecuencia acumulada (F, Fa o Fi)	Frecuencia relativa (fr)	Frecuencia relativa acumulada (Fr)
Es cada intervalo en que se ha decido agrupar los subgrupos de datos	Es el punto medio de cada intervalo	Es el número de veces que se repite un valor o evento	Es el resultado de ir sumando las frecuencias absolutas	Es el cociente entre la frecuencia absoluta y el número de datos	Es el resultado de ir sumando las frecuencias relativas

Ejemplo:

Las edades de pacientes que acuden a un centro de salud en un día, son las siguientes:

2	7	10	16	19
22	6	25	5	20
13	32	13	29	18
20	13	6	12	35

Solución

1. **Hallamos el rango R**

$R = 35 - 2$

$R = 33$

2. **Hallamos el número de intervalos K**

$K = 1 + 3{,}322 * \log(N)$

$K = 1 + 3{,}322 * \log(20)$

$k = 1 + 3{,}322 * (1{,}301)$

$K = 1 + 4{,}323$

$K = 5{,}323$

$K \cong 5$ **se redondea a 5**

3. **Determinamos la amplitud de clase (A)**

$$A = \frac{R}{k}$$

$$A = \frac{33}{5}$$

$A = 6{,}6$

A ≅ 7 **se redondea a 7**

4. Construimos la tabla y realizamos los cálculos:

Formamos los intervalos:

Límite inferior	+ Amplitud	Límite superior
2	+7	9
9	+7	16
16	+7	23
23	+7	30
30	+7	37

5. Calculamos las marcas de clase

$$X = \frac{Ls + Li}{2}$$

$$X = \frac{9+2}{2} = \frac{11}{2} = 5{,}5$$

$$X = \frac{16+9}{2} = \frac{25}{2} = 12{,}5$$

$$X = \frac{23+16}{2} = \frac{39}{2} = 19{,}5$$

$$X = \frac{30+23}{2} = \frac{53}{2} = 26{,}5$$

$$X = \frac{37+30}{2} = \frac{67}{2} = 33{,}5$$

6. Completamos la tabla y seguimos con los otros cálculos de la misma manera que se realizó en el ejemplo de la distribución para datos no agrupados.

Para la frecuencia absoluta, se debe observar los datos que están dentro del intervalo tomando desde el valor del límite inferior hasta antes del límite superior, si hay un número que coincida con el límite superior no se le tomará en cuenta en ese intervalo sino en el siguiente.

Edades (intervalos)	Marcas de clase X	Frecuencia absoluta f o fi	Frecuencia acumulada F	Frecuencia relativa fr	Frecuencia relativa acumulada Fr
[2 - 9)	5,5	5	5	5/20 = 0,25	0,25
[9 - 16)	12,5	5	5 + 5 = 10	5/20 = 0,25	0,25 + 0,25 = 0,5
[16 - 23)	19,5	6	10 + 6 = 16	6/20 = 0,3	0,5 + 0,3 = 0,8
[23 - 30)	26,5	2	16 + 2 - 18	2/20 = 0,1	0,8 + 0,1 = 0,9
[30 - 37)	33,5	2	18 + 2 = 20	2/20 = 0,1	0,9 + 0,1 = 1
Total		20		1	

Interpretación de algunos datos de la tabla:

El valor **f=6** significa que 6 pacientes fueron atendidos cuyas edades variaban entre los 16 y 22 *años.*

El valor **X=12,5** significa que 5 pacientes atendidos tenían entre 9 y 15 años.

El valor **F=18** significa que 18 pacientes fueron atendidos entre los 2 y 29 *años*.

El valor de **fr= 0,3 o 30%** significa que el 0,3 o 30% de los pacientes tenían entre 16 y 22 años

El valor **Fr=0,9 o 90%** significa que el 0,9 o 90% de los pacientes atendidos tenían entre 2 y 29 años.

GRÁFICOS ESTADÍSTICOS

Los gráficos estadísticos son herramientas visuales clave utilizados en el campo de la estadística para representar información de manera clara y comprensible, estos permiten visualizar patrones, tendencias y relaciones en los datos, lo que facilita la interpretación y comunicación de resultados, entre los tipos más comunes de gráficos estadísticos se encuentran los histogramas, gráficos de barras, gráficos de dispersión, gráficos circulares (o de sectores), y gráficos de líneas, cada uno de estos es adecuado para diferentes tipos de datos y objetivos de análisis (Chance, 2002).

Los histogramas son gráficos utilizados para representar la distribución de frecuencias de datos numéricos continuos. Consisten en barras adyacentes cuyas alturas representan las frecuencias de intervalos o clases específicas de datos. Los gráficos de barras, por otro lado, son útiles para representar la distribución de frecuencias de datos categóricos o discretos. Cada barra en un gráfico de barras representa una categoría distinta y la altura de la barra indica la frecuencia o conteo de esa categoría (López & Gómez, 2023).

Los gráficos de dispersión son útiles para visualizar la relación entre dos variables numéricas. Cada punto en el gráfico representa una observación que tiene valores específicos para ambas variables, y la dispersión de los puntos revela la naturaleza de la relación entre las variables (por ejemplo, si existe una correlación positiva, negativa o ninguna). Por otro lado, los gráficos circulares son útiles para mostrar proporciones o

porcentajes de diferentes categorías en un conjunto de datos. Cada sector del círculo representa una categoría específica y su tamaño relativo muestra la proporción que representa dentro del total (Karen et al., 2024).

DIAGRAMA DE BARRAS

El diagrama de barras es una herramienta visual utilizada en estadística para representar datos categóricos o discretos de manera clara y comprensible, este consiste en barras rectangulares colocadas horizontal o verticalmente, donde la longitud de cada una de estas representa la frecuencia o el conteo de una categoría específica; cada barra está separada y etiquetada según la categoría que representa, lo que facilita la comparación entre diferentes condiciones y la identificación de las más frecuentes o relevantes en un conjunto de datos; este diagrama es especialmente útil para resaltar diferencias o patrones entre grupos de datos, ya que permite una visualización directa y fácil de interpretar (Monsalve, 2024).

CARACTERÍSTICA DE UNA GRÁFICA DE BARRAS

Para crear un diagrama de barras efectivo, hay que seguir algunos principios de diseño, por ejemplo, asegurarse de que las barras sean proporcionales a las frecuencias que representan y que estén etiquetadas correctamente para que sean identificables, la elección entre un diagrama horizontal o vertical depende del tipo de datos y la presentación deseada (Aquije et al., 2023).

Ejemplo:

En el paralelo A de Educación Inicial 1 de una escuelita, se les preguntó a los niños sobre su color favorito obteniendo el siguiente resultado, como se ilustra en la tabla siguiente:

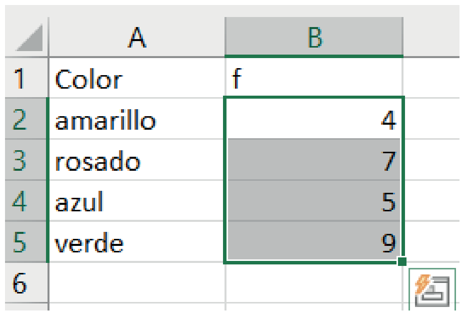
Color	Frecuencia absoluta	Frecuencia relativa	Frecuencia porcentual
Amarillo	4	0,16	16%
Rosado	7	0,28	28%
Azul	5	0,2	20%
Verde	9	0,36	36%
Total	25		

Para graficar el diagrama de barra, se utiliza el programa Excel

- Primero, ingresamos a Excel
- Segundo, colocamos la columna de la variable color y de la frecuencia absoluta
- Tercero, se debe seleccionar la columna de datos de la frecuencia desde B2 – B5

Figura 4

Variable y su frecuencia



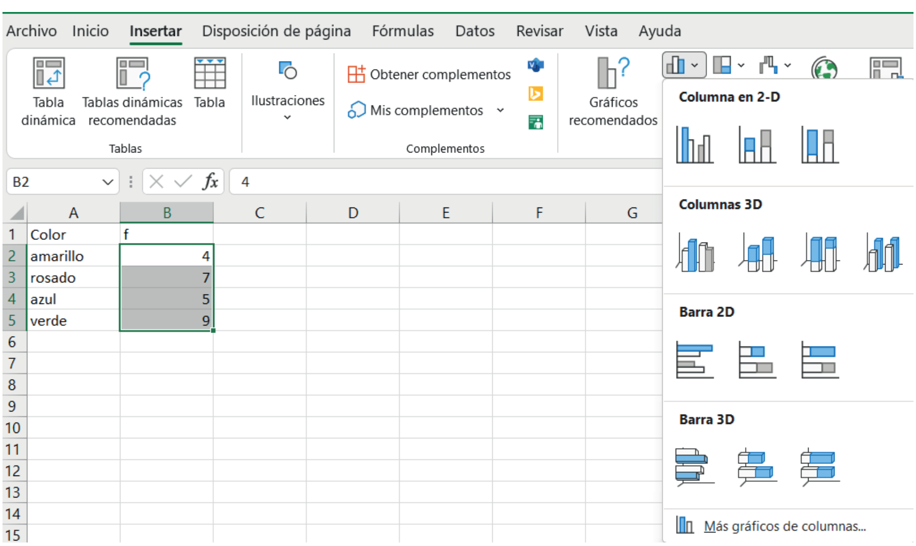
	A	B
1	Color	f
2	amarillo	4
3	rosado	7
4	azul	5
5	verde	9
6		

Nota: Esta figura muestra una distribución de la variable color junto a su frecuencia absoluta misma que fue elaborada por los autores, 2024

- Cuarto, mantenemos la selección y vamos al icono **insertar** y seleccionamos el diagrama de barras

Figura 5

Generar un diagrama de barras en Excel

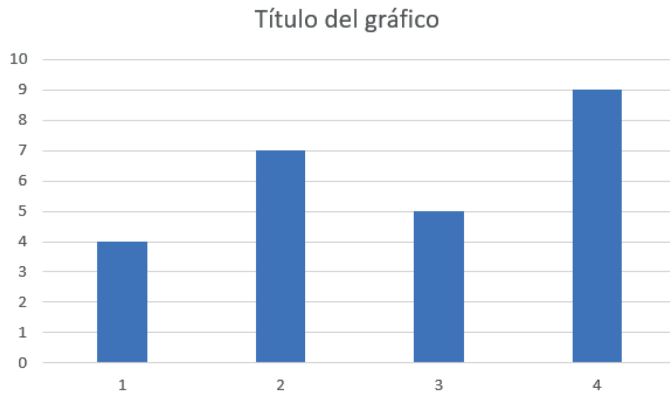


Nota: Esta figura representa la creación de un gráfico estadístico, mismo que fue elaborada por los autores, 2024

- Quinto, al hacer clic, inmediatamente aparece la gráfica

Figura 6

Gráfico estadístico



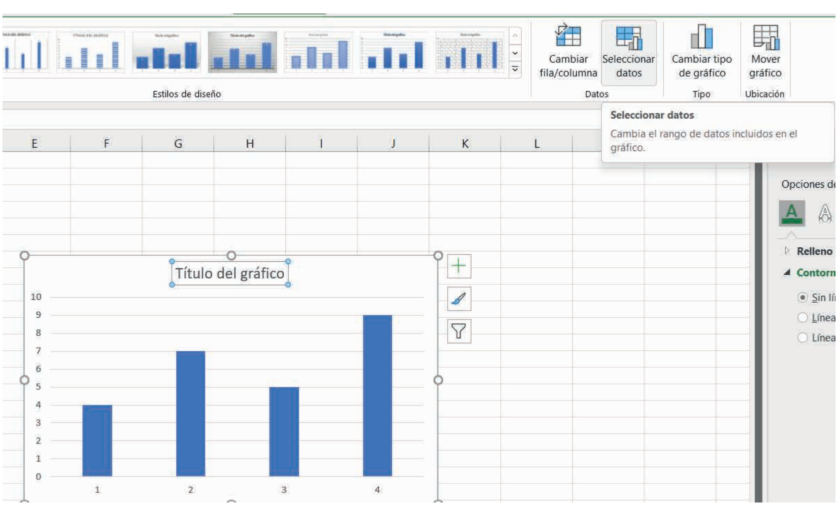
Nota: Esta figura representa un diagrama de barras para frecuencia absoluta, misma que fue elaborada por los autores, 2024

- Sexto, ahora vamos a configurar el gráfico, para que en el eje horizontal aparezcan las etiquetas de los colores

  Para ello hacemos doble clic sobre el diagrama de barras y seleccionamos la opción en el ícono de **seleccionar datos**

Figura 7

Edición de título para el diagrama



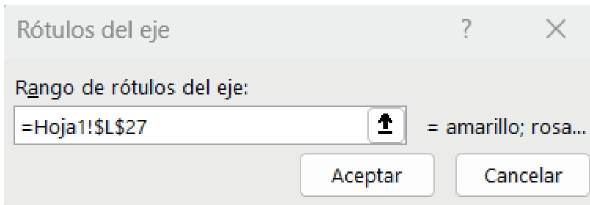
Nota: Esta figura representa la configuración del título del diagrama de barras, misma que fue elaborada por los autores, 2024

- Séptimo, aparece una nueva ventana, para editar el eje horizontal se debe hacer clic en donde dice **etiquetas del eje horizontal (categorías)**

Clic en **editar** y marcar la columna A2 – A5 en nuestros datos de **color**

Figura 8

Edición de ejes



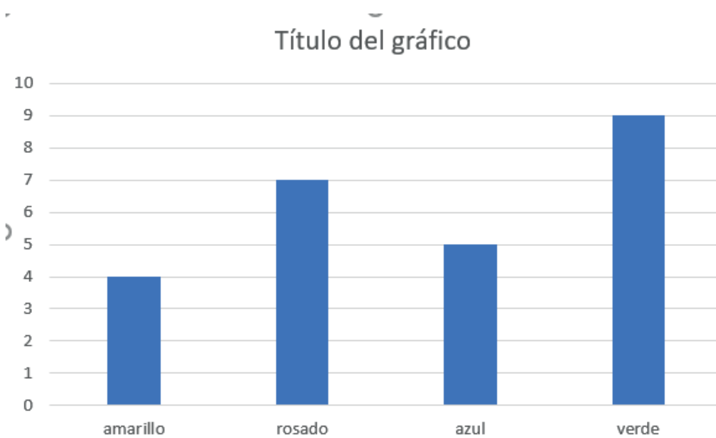
Nota: Esta figura muestra inserción de las categorías de la variable color en el eje horizontal del diagrama de barras, misma que fue elaborada por los autores, 2024

Dar en clic en **aceptar**

Se observan ya los colores con su respectiva frecuencia

Figura 9

Eje horizontal y sus categorías



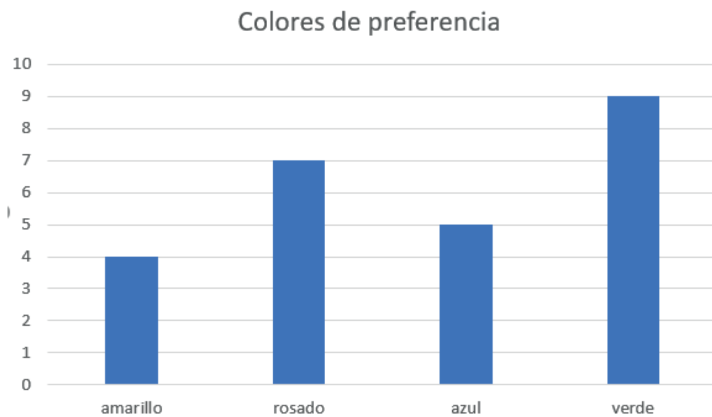
Nota: Esta figura muestra lo que representa cada barra, misma que fue elaborada por los autores, 2024

- Octavo, ahora se va a editar el nombre de la gráfica y cambiar colores si se desea

Hacer doble clic en **Título del gráfico** y cambiar el nombre

Figura 10

Título principal

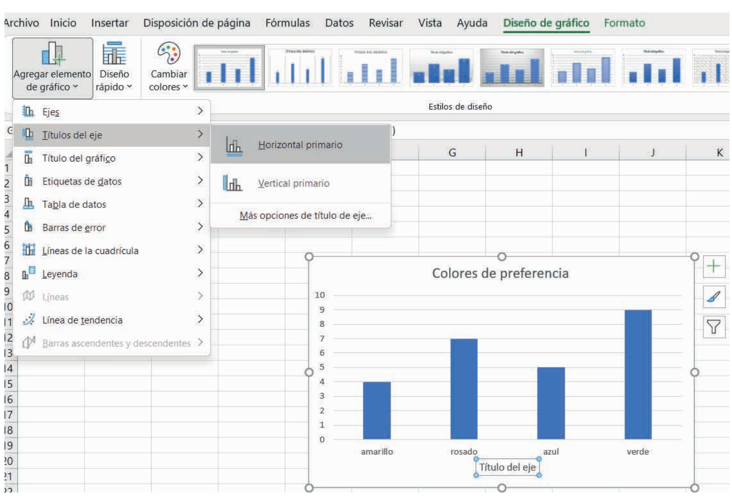


Nota: Esta figura muestra el título principal cambiado por “colores de preferencia” misma que fue elaborada por los autores, 2024

- Noveno, hacer clic en **Diseño de gráfico,** posterior clic en **Agregar elemento de gráfico** y se va a insertar los títulos de los ejes horizontal (colores) y vertical (frecuencia absoluta)

Figura 11

Inserción de título para el eje vertical

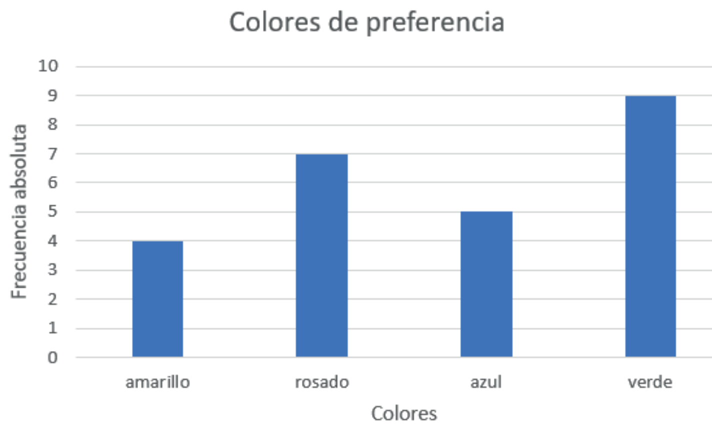


Nota: Esta figura como agregar un elemento que represente el título del eje vertical de la gráfica, misma que fue elaborada por los autores, 2024

Editamos los nombres de los ejes

Figura 12

Título del eje vertical

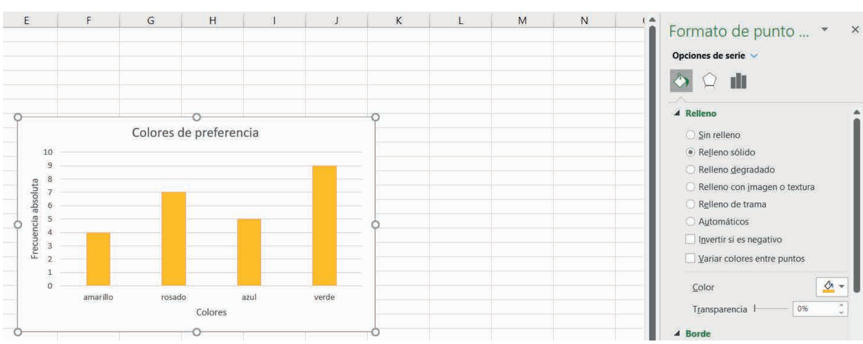


Nota: Esta figura muestra los títulos agregados, misma que fue elaborada por los autores, 2024

- Décimo, para cambiar el color de las barras, hacer clic sobre una de las barras y aparece una sub-ventana **(Formato de serie)** en la parte derecha y ahí cambiamos el color en el ícono **relleno** se van cambiando lo colores de su preferencia.

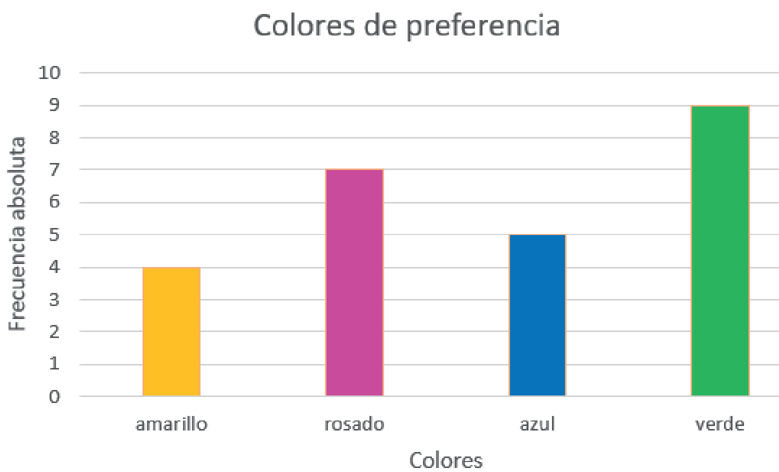
Figura 13

Cambio de color



Nota: Esta figura muestra el cambio del color de las barras, misma que fue elaborada por los autores, 2024

Se ha cambiado los colores barra por barra

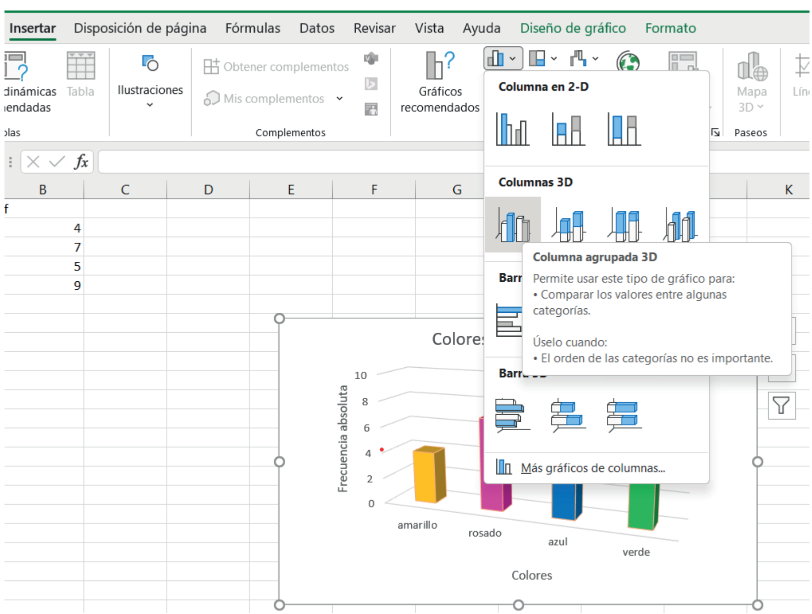


Nota: Esta figura muestra la edición en el cambio de color de cada barra, misma que fue elaborada por los autores, 2024

Finalmente, si se desea un gráfico en 3D, se hace clic en la pestaña insertar y luego en el ícono de gráficos de barras y se selecciona el estilo **Columnas 3D**

Figura 14

Presentación de otro estilo de barras

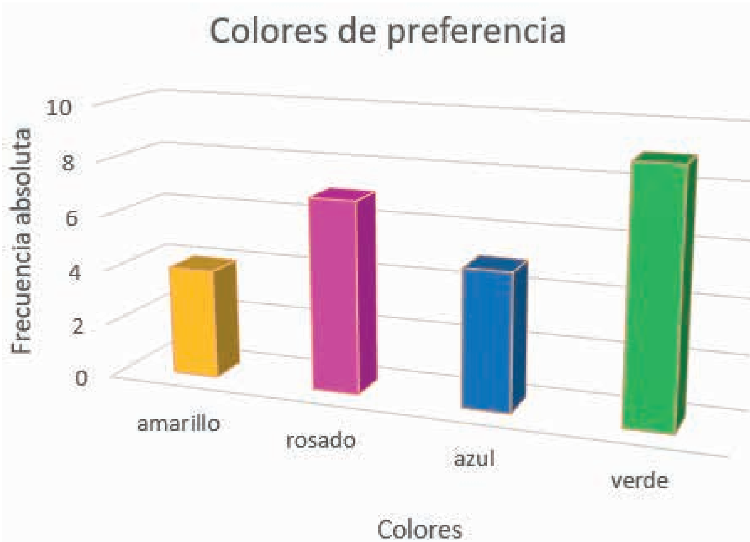


Nota: Esta figura representa un diagrama de barras en 3D, misma que fue elaborada por los autores, 2024

Resultado de la gráfica

Figura 15

Diagrama 3D



Nota: Esta figura muestra un diagrama de barras en 3 dimensiones, misma que fue elaborada por los autores, 2024

HISTOGRAMAS

Un histograma es un tipo de gráfico utilizado en estadística para representar la distribución de frecuencias de datos numéricos continuos, consiste en barras adyacentes donde la base de cada una de estas representa un intervalo de valores y la altura de la barra representa la frecuencia o cantidad de observaciones dentro de ese intervalo, estos son útiles para visualizar la forma y la dispersión de los datos, así como para identificar patrones o características importantes en la distribución, por ejemplo, un histograma puede mostrar si los datos siguen una distribución normal, están sesgados hacia un lado o muestran múltiples picos (Guttman, 1979).

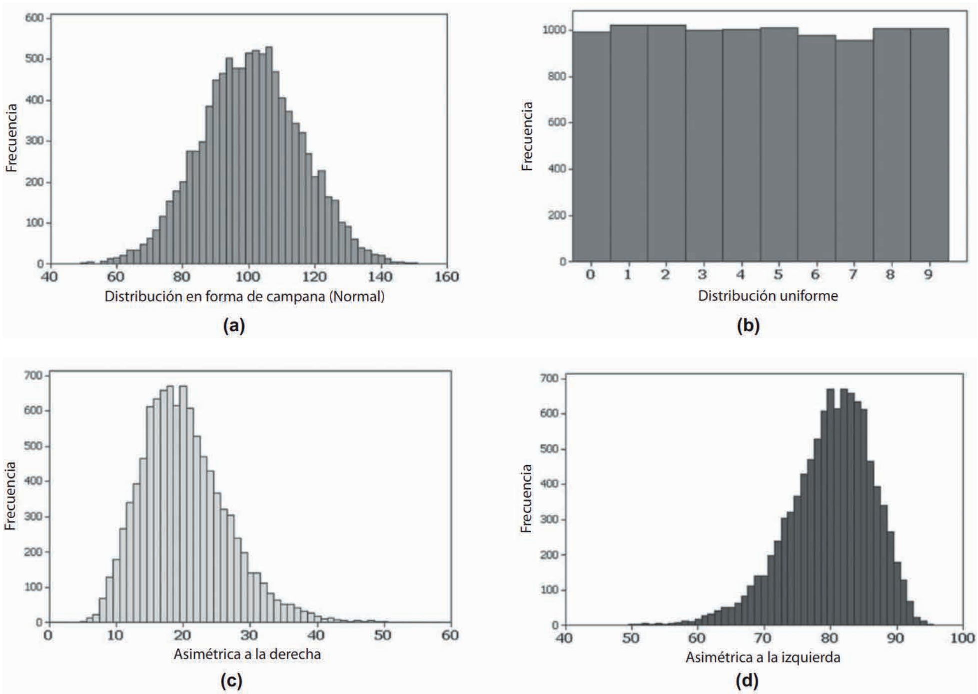
Para construir un histograma efectivo, hay que seleccionar el número adecuado de intervalos (o clases) y el ancho de cada uno, de manera que se represente la distribución de los datos sin perder detalles importantes, además, hay que etiquetar adecuadamente los ejes del histograma para indicar qué valores representan las barras horizontales y verticales.

Estos gráficos se utilizan en diversas áreas como la investigación científica, la ingeniería, la economía y la salud, ya que dan una representación visual poderosa de la estructura y las características de los conjuntos de datos numéricos continuos, facilitando la interpretación y el análisis estadístico de estos (Rodríguez-Alveal et al., 2024).

Formas comunes de Distribución

Figura 16

Formas comunes de Distribución



Nota: Esta figura fue tomada del Libro Estadística Mario F. Triola, pg. 52

Ejemplo:

En la siguiente gráfica se muestra una distribución de datos sobre las edades de las personas que ingresan a Laguna Mall el día domingo de 9-10 am en promedio. Elaborar:

Un histograma de frecuencia absoluta

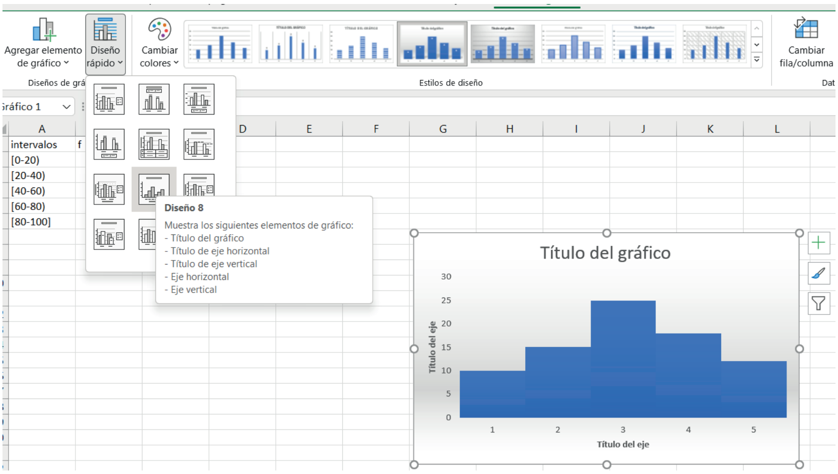
Edad (intervalo)	**F**	**Fr**	**fr %**
[0-20)	10	0,125	12,5%
[20-40)	15	0,1875	18,75%
[40-60)	25	0,3125	31,25%
[60-80)	18	0,225	22,5%
[80-100]	12	0,15	15%
Total	80	1	100%

Solución para el Histograma de frecuencia absoluta

- Primero, ingresar a Excel y seleccionar la columna desde B2-B6 luego seleccionar el diagrama de barra y en el ícono **Diseño rápido** seleccionar el Diseño 8.

Figura 17

Gráfico estadístico



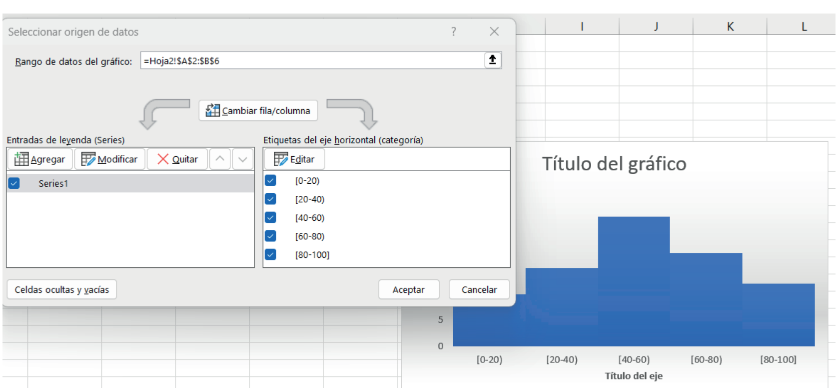
Nota: Esta figura representa un Histograma de distribución de frecuencias absolutas, misma que fue elaborada por el autor, 2024

- Segundo, procedemos a editar las etiquetas horizontales, verticales y el título

  Hacer clic en Diseño de tabla y luego en **Etiquetas del eje horizontal** y seleccionamos desde A2-A6 que corresponde a los intervalos.

Figura 18

Histograma

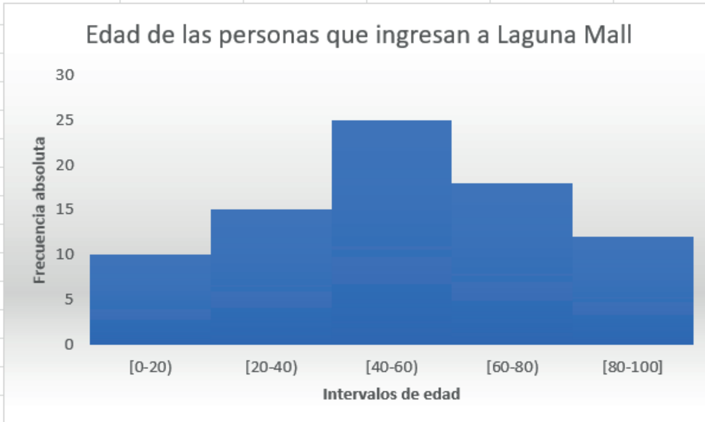


Nota: Esta figura muestra la edición del eje horizontal para el histograma, misma que fue elaborada por el autor, 2024

- Tercero, colocamos los nombres de los ejes y título

Figura 19

Histograma con títulos



Nota: Esta figura muestra un histograma con los títulos principal, horizontal y vertical, misma que fue elaborada por el autor, 2024

POLÍGONO DE FRECUENCIAS

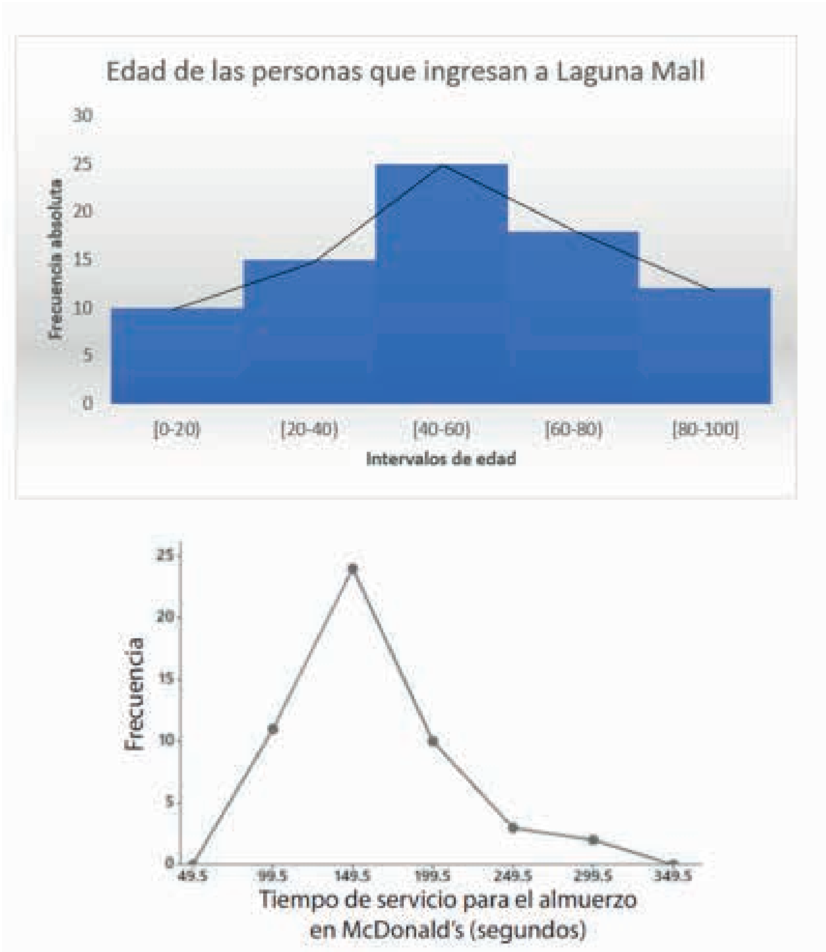
El polígono de frecuencias es una representación gráfica utilizada en estadística para visualizar la distribución de frecuencias de un conjunto de datos discretos, traza puntos en un plano cartesiano, donde cada punto representa un valor específico de la variable y su frecuencia asociada en el conjunto de datos, posterior, se unen con líneas rectas para formar un polígono que representa la distribución de frecuencias en diferentes valores (Ramírez et al., 2024).

Para crear un polígono de frecuencias, se deben seguir varios pasos.

- Se calculan las frecuencias de cada valor en el conjunto de datos. Luego, se organizan estos valores y sus frecuencias en pares ordenados (x, y), donde x representa el valor y y representa la frecuencia correspondiente.
- Posteriormente, se colocan estos puntos en un sistema de coordenadas cartesianas, etiquetando adecuadamente los ejes para representar los valores y las frecuencias.
- Por último, se unen los puntos con líneas rectas para formar el polígono de frecuencias.

Figura 20

Polígono de frecuencias en un histograma



Nota: La gráfica 2, representa un Polígono de Frecuencias

Ejemplo:

Las calificaciones de los exámenes de parcial de Estadística de los estudiantes de Educación Inicial obtenidas sobre 20 puntos, se muestran en la siguiente tabla de distribución, con estos datos construir el Polígono de frecuencias.

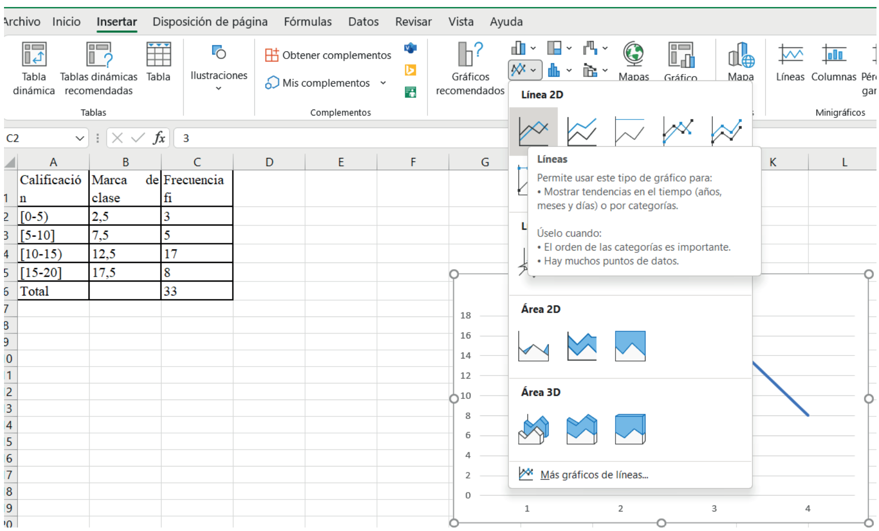
Calificación	Marca de clase	Frecuencia fi
[0-5)	2,5	3
[5-10]	7,5	5
[10-15)	12,5	17
[15-20]	17,5	8
Total		33

Solución:

- Primero, ingresar en Excel los datos en columnas tanto la calificación, la marca de clase como la frecuencia
- Segundo, seleccionar los datos de la columna **frecuencia** e insertar un gráfica en este caso seleccionamos **insertar líneas 2D**

Figura 21

Polígono de frecuencias

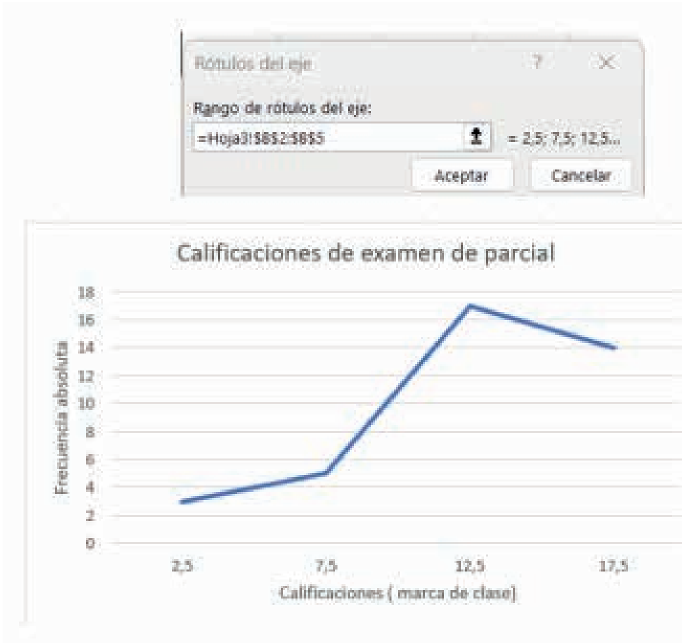


Nota: Esta figura muestra la construcción de un Polígono de frecuencias en Excel, misma que fue elaborada por los autores, 2024

- Tercero, editamos la gráfica colocando en eje horizontal. Seleccionar la columna de las **marcas de clase** y hacer clic en **Aceptar y colocamos los títulos en los ejes y nombre de la gráfica.**

Figura 22

Polígono de frecuencias



Nota: Esta figura muestra la edición de los datos del eje horizontal, misma que fue elaborada por los autores, 2024

PASTEL O DIAGRAMA CIRCULAR

Los diagramas circulares, también conocidos como gráficos de pastel, son una herramienta visual comúnmente utilizada en estadística y presentaciones de datos para representar proporciones o porcentajes de diferentes categorías dentro de un conjunto de datos, los mismos se basan en un círculo dividido en sectores, donde cada sector representa una categoría específica y su tamaño angular es proporcional al porcentaje que representa en el total, la suma de todos los sectores completa el círculo, lo que permite visualizar de manera clara la distribución relativa de las categorías (Ghahramani, 2015).

Los diagramas circulares son especialmente útiles cuando se desea destacar la proporción o contribución de cada categoría dentro de un conjunto de datos; sin embargo, es importante utilizarlos con precaución, ya que pueden volverse menos efectivos cuando se intenta representar demasiadas categorías o cuando las diferencias entre las proporciones son mínimas (Giuliana & Silvana, 2023).

Para crear un diagrama circular efectivo, es esencial etiquetar cada sector con el nombre de la categoría correspondiente y su porcentaje o proporción relativa, es recomendable ordenar los sectores de manera que sea fácil identificar las categorías más grandes y más pequeñas, los diagramas circulares son ampliamente utilizados en informes,

presentaciones y medios visuales debido a su capacidad para resaltar de manera intuitiva las proporciones y relaciones entre diferentes partes de un conjunto de datos, lo que facilita la comunicación y comprensión de la información estadística (Aguilar et al., 2024).

Ejemplo:

Con los datos de la tabla siguiente, correspondiente al tiempo que se demoran unas partículas en dar una vuelta elíptica, elaborar un diagrama circular con las frecuencias porcentuales.

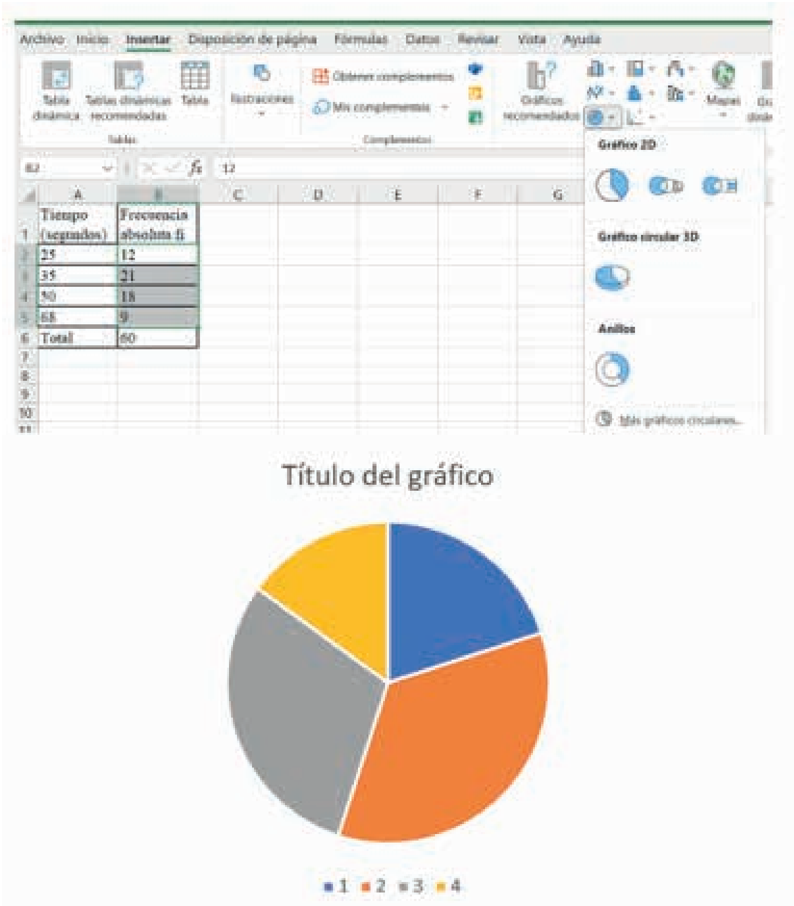
Tiempo (segundos)	**Frecuencia absoluta fi**	**Frecuencia relativa fr**	**Frecuencia porcentual fr%**
25	12	0,2	20%
35	21	0,35	35%
50	18	0,3	30%
68	9	0,15	15%
Total	60	1	100%

Solución:

- Primero, ingresar a Excel y colocar los datos en las columnas
- Segundo, seleccionamos los datos de la columna de frecuencia absoluta desde B2-B5, clic en el ícono **insertar** y seleccionar **Gráfico 2D**

Figura 23

Diagrama estadístico circular

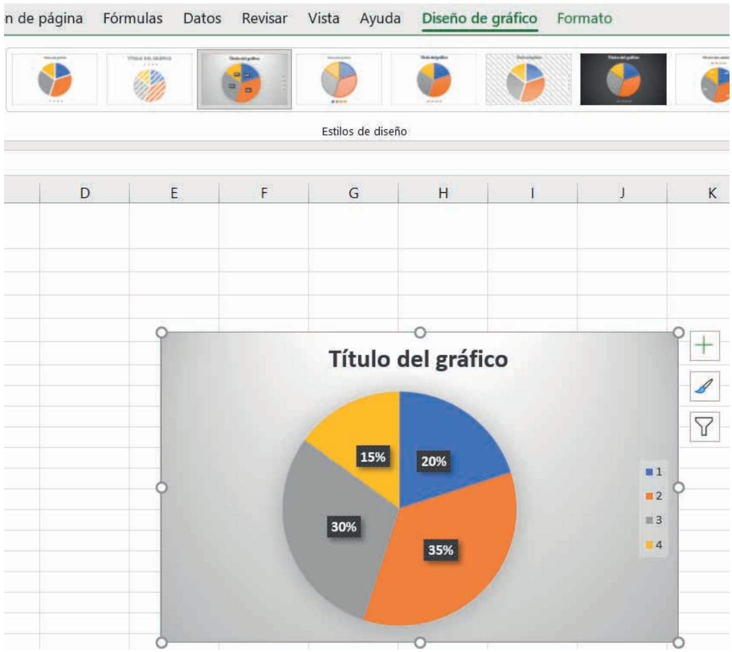


Nota: Esta figura representa un diagrama circular para una distribución de frecuencias, misma que fue elaborada por los autores, 2024

- Tercero, cambiamos estilo de la imagen para que aparezcan los porcentajes (frecuencia relativa porcentual), para ello hacer clic en la gráfica, luego clic en **Diseño de gráfico** y seleccionamos el **estilo 3.**

Figura 24

Edición del Diagrama circular

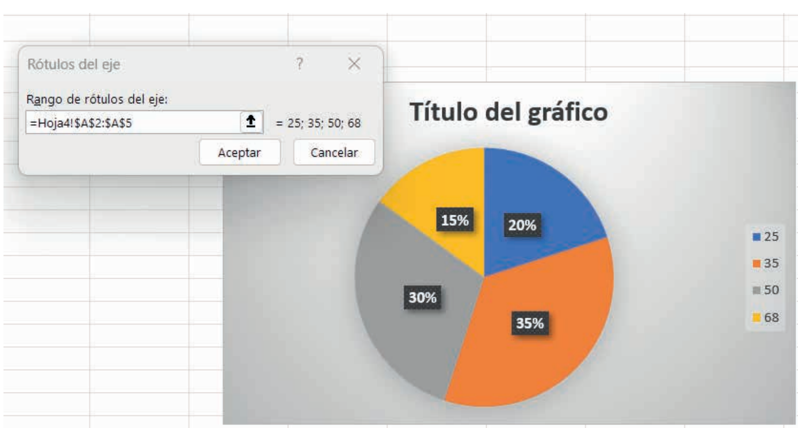


Nota: Esta figura representa un diagrama circular con porcentajes de frecuencia absoluta, misma que fue elaborada por los autores, 2024

- Cuarto, hacer clic en la imagen, luego en **Seleccionar datos** y editamos los datos del eje horizontal, seleccionamos los datos de A2-A5 y clic en **Aceptar.**

Figura 25

Diagrama circular



Nota: Esta figura muestra la edición del título del gráfico, misma que fue elaborada por los autores, 2024

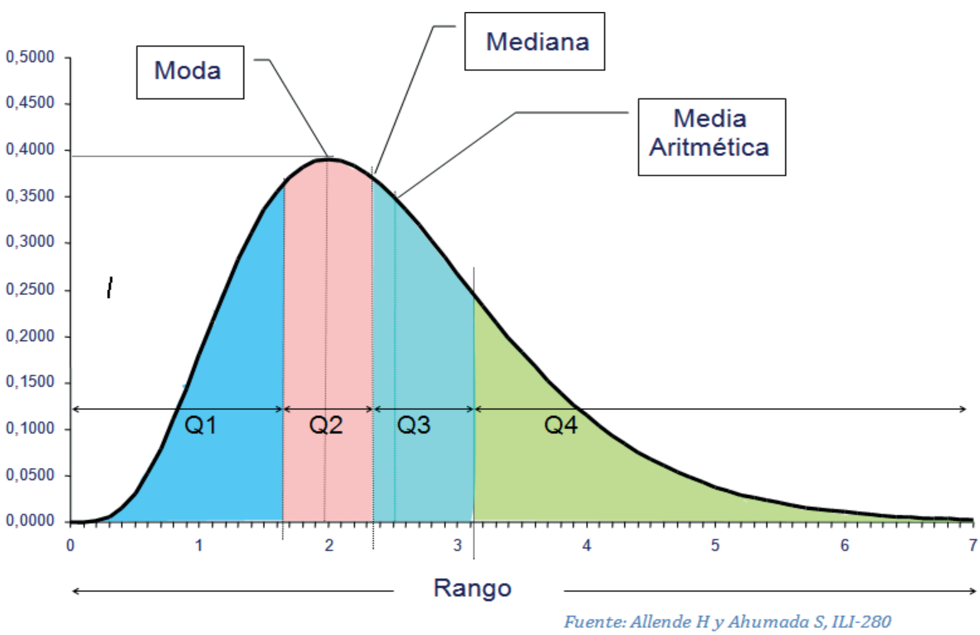
- Quinto, finalmente, colocamos el nombre

Figura 26

Diagrama circular



Nota: Esta figura muestra los título y colores de representación de sus frecuencias absolutas, misma que fue elaborada por los autores, 2024



UNIDAD 2

MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN CONTENIDO

¿QUÉ SON LAS MEDIDAS DE TENDENCIA CENTRAL?

Las medidas de tendencia central son estadísticas para describir la ubicación central o típica de un conjunto de datos, las tres principales son la media, la mediana y la moda; la media aritmética es la medida más común y se calcula sumando valores en el conjunto de datos y dividiéndolos por el número de observaciones, sensible a los valores atípicos y verse afectada por distribuciones sesgadas (Millán, 2023).

La mediana es el valor que se encuentra en el centro de un conjunto de datos ordenados de menor a mayor, esta es menos sensible a los valores extremos y proporciona una medida robusta de la ubicación central.

La moda es el valor que aparece con mayor frecuencia en un conjunto de datos y puede ser útil para describir distribuciones con picos o agrupamientos de datos.

El uso adecuado de medidas de tendencia central depende del tipo de datos y del objetivo del análisis, la media es útil cuando se desea una representación promedio de los datos y cuando la distribución es simétrica y no está sesgada.

La mediana es preferible en presencia de valores extremos o cuando se necesita una medida más robusta contra la variabilidad en los datos.

La moda es efectiva para identificar los valores más comunes en un conjunto de datos y puede ser útil en estudios de preferencias o frecuencias; en conjunto, estas medidas proporcionan información valiosa sobre la ubicación central de los datos y ayudan a resumir la distribución de manera concisa y significativa para el análisis estadístico (Cujba & Pifarré, 2024).

MEDIA, MEDIANA Y MODA PARA DATOS NO AGRUPADOS

Figura 28

Modelo de una matriz de responsabilidad RACI



Nota: Esta figura representa un ejemplo de una matriz de responsabilidad RACI, misma que fue elaborada por los autores, 2024

Ejemplos:

- Encuentre el **promedio o media** de los datos: 12 , 15 , 18 , 20

$$\bar{x} = \frac{\sum x}{N}$$

$$\bar{x} = \frac{12 + 15 + 18 + 20}{4}$$

$$\bar{x} = 16{,}25$$

- Encuentre la **mediana** de los datos: 5, 7, 4, 3, 8, 7

  Ordenar los datos y tomarlos una sola vez los datos repetidos: 3, 4, 5, 7, 8

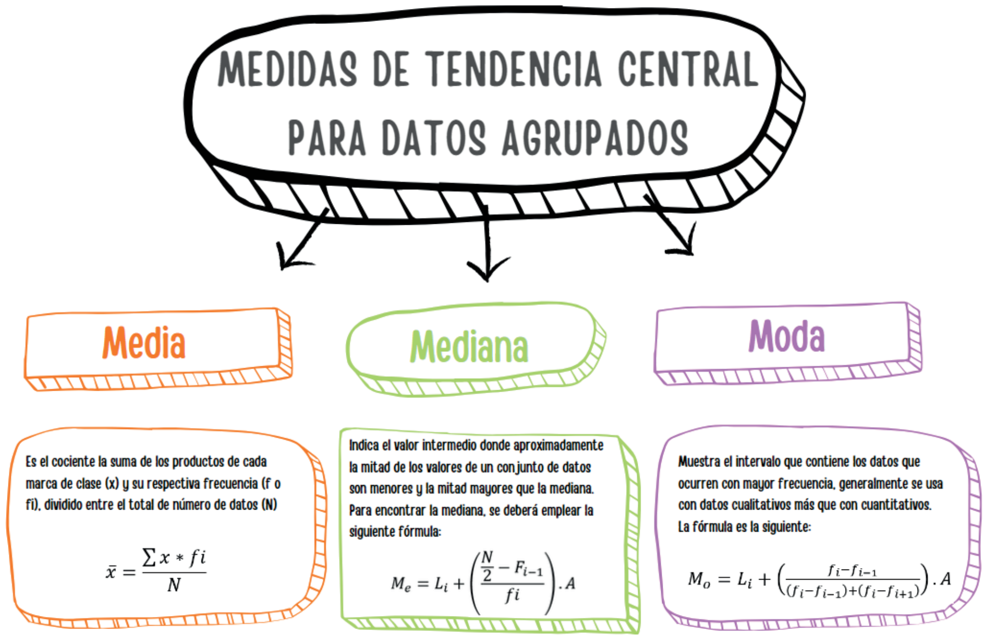
- Encuentre la **moda** de los datos: 2, 3, 3, 4, 5, 5, 5

  Se observa que el dato que más aparece es el número 5

MEDIA, MEDIANA Y MODA PARA DATOS AGRUPADOS

Figura 29

Modelo de una matriz de responsabilidad RACI



Nota: Esta figura representa un ejemplo de una matriz de responsabilidad RACI, misma que fue elaborada por los autores, 2024

EXPLICACIÓN DE LAS FÓRMULAS

Media

$$\bar{x} = \frac{\sum x * fi}{N}$$

Donde:

$x * f_i$ es el producto de las marcas de clase por su frecuencia absoluta

N es el número total de datos

Mediana

$$M_e = L_i + \left(\frac{\frac{N}{2} - F_{i-1}}{fi}\right).A$$

Nota: Para localizar el intervalo donde se encuentra la mediana se debe tener en cuenta lo siguiente:

$$posición\ intervalo = \frac{N}{2} \quad si\ N\ es\ par$$

$$posición\ intervalo = \frac{N+1}{2} \quad si\ N\ es\ impar$$

Finalmente, se ubica el intervalo de clase donde se encuentra la mediana, en el primer intervalo de clase donde la **frecuencia acumulada (Fi o Fa)** es igual o mayor que la posición encontrada por las fórmulas.

Donde:

L_i Límite inferior del intervalo en el cual se encuentra la mediana

N Número total de datos

F_{i-1} Frecuencia acumulada del intervalo anterior al que se encuentra la mediana

A Amplitud del intervalo donde se encuentra la mediana

f_i Frecuencia absoluta del intervalo en el que se encuentra la mediana

Moda

$$M_o = L_i + \left(\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right).A$$

Nota: el intervalo modal se encuentra en el intervalo que contiene la mayor frecuencia absoluta.

Donde

L_i Límite inferior del intervalo en el que se encuentra la moda

f_i Frecuencia absoluta del intervalo en que se encuentra la moda

f_{i-1} Frecuencia absoluta del intervalo anterior al de la moda

f_{i+1} Frecuencia absoluta del intervalo siguiente al de la moda

A Amplitud del intervalo donde se encuentra la moda

Ejemplo:

Encuentre la media, mediana y moda de la siguiente distribución de frecuencias:

Intervalos	Marca de clase x	Frecuencia absoluta fi	Frecuencia acumulada Fi
[0 - 4)	2	3	3
[4 - 8)	6	5	8
[8 - 12)	10	6	14
[12, 16)	14	4	18
[16 - 20)	18	3	21
Total		21	

Solución:

Media

Usamos la fórmula:

$$\bar{x} = \frac{\sum x * fi}{N}$$

$$\bar{x} = \frac{x_1.f_1 + x_2.f_2 + x_3.f_3 + x_4.f_4 + x_5.f_5}{N}$$

$$\bar{x} = \frac{2(3) + 6(5) + 10(6) + 14(4) + 18(3)}{21}$$

$$\bar{x} = \frac{206}{21} = 9{,}81$$

Mediana

Primero, localizamos el intervalo donde se encuentra la mediana, como N= 21 es impar, aplicamos la fórmula:

$$\textit{posición intervalo} = \frac{N+1}{2}$$

$$\textit{posición intervalo} = \frac{21+1}{2} = \frac{22}{2} = 11$$

Buscamos el valor en la columna de la Frecuencia acumulada **F** cuyo valor sea mayor o igual a **11**

Figura 30

Distribución de frecuencias

Intervalos	Marca de clase x	Frecuencia absoluta fi	Frecuencia acumulada Fi
$[0-4)$	2	3	3
$[4-8)$	6	5	8
$[8-12)$	10	6	14
$[12,16)$	14	4	18
$[16-20)$	18	3	21
Total		21	

→ F_{i-1}

Nota: Esta figura muestra la posición de la mediana, así como su frecuencia acumulada anterior, misma que fue elaborada por el autor, 2024

Segundo, aplicamos la fórmula:

$$M_e = L_i + \left(\frac{\frac{N}{2} - F_{i-1}}{fi}\right).A$$

$$M_e = 8 + \left(\frac{\frac{21}{2} - 8}{6}\right).4$$

$$M_e = 8 + \left(\frac{10,5 - 8}{6}\right).4$$

$$M_e = 9,667$$

Moda

Primero, encontramos el intervalo modal que es el intervalo donde se tiene la mayor frecuencia absoluta.

Figura 27

Tabla de Distribución de frecuencias

Intervalos	**Marca de clase x**	**Frecuencia absoluta fi**	**Frecuencia acumulada Fi**
$[0-4)$	2	3	3
$[4-8)$	6	5 → f_{i-1}	8
$[8-12)$	10	**6** f_i	14
$[12,16)$	14	4 → f_{i+1}	18
$[16-20)$	18	3	21
Total		21	

Nota: Esta figura muestra la posición de la moda, así como su frecuencia absoluta anterior y posterior, misma que fue elaborada por los autores, 2024

Segundo, aplicamos la fórmula:

$$M_o = L_i + \left(\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right).A$$

$$M_o = 8 + \left(\frac{6-5}{(6-5)+(6-4)}\right).4$$

$$M_o = 9{,}333$$

¿QUÉ SON LAS MEDIDAS DE DISPERSIÓN?

Las medidas de Dispersión o Variabilidad, describen cómo de extendidos (separados) o agrupados (juntos) están los datos en un conjunto. En otras palabras, nos indican el grado de variabilidad que existe entre los valores de una variable alrededor de una medida de posición central. Es decir, las medidas de dispersión nos permiten saber si los datos se encuentran estrechamente agrupados, si se encuentran ampliamente dispersos o si son iguales (Rodríguez et al., 2023).

Rango, desviación media, varianza, desviación estándar y coeficiente de variaza para datos no agrupados y agrupados

Las medidas de Dispersión o Variabilidad, describen cómo de extendidos (separados) o agrupados (juntos) están los datos en un conjunto, en otras palabras, nos indican el grado de variabilidad que existe entre los valores de una variable alrededor de una medida de posición central. Las medidas de dispersión permiten saber si los datos se encuentran estrechamente agrupados, ampliamente dispersos o iguales (Dexeus, 2001).

Rango

El rango o recorrido es la distancia que existe entre los valores extremos máximo y mínimo de un conjunto de datos ordenados, es decir es la diferencia entre el valor máximo de datos y el valor mínimo de datos.

Rango = (valor máximo de datos) 2 (valor mínimo de datos)

Propiedad importante del rango

El rango utiliza sólo los valores máximo y mínimo de los datos, por lo que es muy sensible a los valores extremos. El rango no es resistente.

Como el rango usa solo los valores máximo y mínimo, no toma en cuenta todos los valores y no refleja la variación entre los valores de los datos.

Ejemplo:

Encuentre el rango de las velocidades de descarga de datos (en Mbps) para la empresa de servicio de internet Redecom:

38,5 55,6 22,4 14,1 23,1

Solución:

El rango se encuentra restando del valor mayor el valor menor

Rango = (valor máximo) - (valor mínimo)

$$R = 55{,}6 - 14.1 Mbps$$

$$R = 41{,}5 Mbps$$

DESVIACIÓN MEDIA

Es la media de las distancias de los datos a la meda de los datos de los que dispongamos. Es una medida robusta, fácil de interpretar y útil para comparar la variabilidad de conjuntos de datos con las mismas unidades.

Propiedades importantes de la desviación media:

1. No es afectada por los valores extremos, es decir no se ve afectada por la presencia de valores extremos en el conjunto de datos, a diferencia de la varianza y la desviación estándar, que sí son sensibles a estos valores.
2. Se expresa en las mismas unidades que los datos originales, lo que facilita su interpretación en términos prácticos. Por ejemplo, si la desviación media de las calificaciones de un examen es de 5 puntos, significa que, en promedio, las calificaciones se desviaron 5 puntos de la media.
3. La desviación media siempre es igual a cero cuando todos los valores del conjunto de datos son iguales a la media aritmética. Esto indica que no hay variabilidad en los datos.

4. Cuanto mayor sea el valor de la desviación media, mayor es la dispersión de los datos.

Fórmula para calcular la desviación media

Ejemplo 1: **Para datos No Agrupados**

Encuentre el valor de la desviación media de la siguiente distribución: 3,4,5,5,5,6,17,9

Solución:

Primero se debe calcular el promedio o media aritmética

$$\bar{x} = \frac{\sum x_i}{N} = \frac{3+4+5+5+5+6+17+9}{8} = \frac{54}{8} = 6{,}75$$

Segundo se calcula la desviación media

$$DM = \frac{\sum|x-\bar{x}|}{N}$$

$$DM = \frac{|3-6{,}75|+|4-6{,}75|+|5-6{,}75|+|5-6{,}75|+|5-6{,}75|+|6-6{,}75|+|17-6{,}75|+|9-6{,}75|}{8}$$

$$DM = \frac{3{,}75+2{,}75+1{,}75+1{,}75+1{,}75+0{,}75+10{,}25+2{,}25}{8}$$

$$DM = \frac{25}{8} = 3{,}125$$

Ejemplo 2: **Para datos Agrupados**

Determine la desviación media para la siguiente distribución de datos, respecto a la edad en años de los estudiantes de Educación Inicial del segundo nivel.

Edad (intervalos)	Frecuencia absoluta	X (marca de clase)	f * X
17 - 20	8	18,5	8*18,5= 148
21 - 24	12	22,5	12*22,5= 270
25 - 28	6	26,5	6*26,5= 159
29 - 32	5	30,5	5*30,5= 152,5
33 - 36	10	34,5	10*34,5= 345

Solución:

Primero se encuentra la media aritmética:

$$\bar{x} = \frac{\sum f.X}{N}$$

$$\bar{x} = \frac{148+270+159+152{,}5+345}{41} = \frac{1074{,}5}{41} = 26{,}2$$

Segundo construimos la siguiente tabla con los cálculos descritos:

Edad (intervalos)	Frecuencia absoluta	X (marca de clase)	$\lvert X - \bar{x} \rvert$	$f * \lvert X - \bar{x} \rvert$
17 - 20	8	18,5	I18,5 - 26,2I = 7,7	8 * 7,7 = 61,6
21 - 24	12	22,5	I22,5 - 26,2I = 3,7	12 * 3,7 = 44,4
25 - 28	6	26,5	I26,5 - 26,2I = 0,3	6 * 0,3 = 1,8
29 - 32	5	30,5	I30,5 - 26,2I = 4,3	5 * 4,3 = 21,5
33 - 36	10	34,5	I34,5 - 26,2I = 8,3	10 * 8,3 = 83

$$DM = \frac{\sum f * |x - \bar{x}|}{N}$$

$$DM = \frac{61{,}6 + 44{,}4 + 1{,}8 + 21{,}5 + 83}{41} = \frac{212{,}3}{41} = 5{,}178$$

Varianza

La varianza ***s***² de un conjunto de valores es una medida de variación, es la media aritmética de los cuadrados de las desviaciones de la media; indica qué tan dispersos están los valores del conjunto de datos respecto a su valor promedio (Pino & Estrella, 2012).

Hay que tener en cuenta que las fórmulas de la varianza y la desviación estándar son diferentes para una muestra que para una población.

Fórmula de la varianza:

Para datos No agrupados No ordenados	
Para una población	**Para una muestra**
$\sigma^2 = \frac{\sum(x - \bar{x})^2}{N}$	$s^2 = \frac{\sum(x - \bar{x})^2}{n - 1}$
Para datos No agrupados Ordenados	
Para una población	**Para una muestra**
$\sigma^2 = \frac{\sum f * (x - \bar{x})^2}{N}$	$s^2 = \frac{\sum f * (x - \bar{x})^2}{n - 1}$
Para datos Agrupados	
Para una población	**Para una muestra**
$\sigma^2 = \frac{\sum f * (X - \bar{x})^2}{N}$	$s^2 = \frac{\sum f * (X - \bar{x})^2}{n - 1}$

Ejemplo:

Determine la varianza de la siguiente distribución:

Edad	fi
13	3
14	5
15	7
16	10
17	11
18	8
19	2
Total	46

- Primero calculamos el promedio o media aritmética:

$$\bar{x} = \frac{13(3) + 14(5) + 15(7) + 16(10) + 17(11) + 18(8) + 19(2)}{46} = \frac{743}{46} = 16{,}15\ años$$

- Segundo completamos la tabla con los cálculos respectivos:

Edad	fi	$x - \bar{x}$	$(x - \bar{x})^2$	$f * (x - \bar{x})^2$
13	3	13 - 16,15 = -3,15	$(-3{,}15)^2$ = 9,92	3 * 9,92 = 29,76
14	5	14 - 16,15 = -2,15	$(-2{,}15)^2$ = 4,62	5 * 4,62 = 23,1
15	7	15 - 16,15 = -1,15	$(-1{,}15)^2$ = 1,32	7 * 1,32 = 9,24
16	10	16 - 16,15 = -0,15	$(-0{,}15)^2$ = 0,02	10 * 0,02 = 0,2
17	11	17 - 16,15 = -0,15	$(0{,}85)^2$ = 0,72	11 * 0,72 = 7,92
18	8	18 - 16,15 = 1,85	$(1{,}85)^2$ = 3,42	8 * 3,42 = 11,42
19	2	19 - 16,15 = 2,85	$(2{,}85)^2$ = 8,12	2 * 8,12 = 10,2
Total	46			

- Tercero, se determina la varianza:

$$\sigma^2 = \frac{\sum f * (x - \bar{x})^2}{N}$$

Tener en cuenta que la sumatoria se debe realizar con todos los datos y sus respectivas frecuencias

$$\sigma^2 = \frac{29{,}76 + 23{,}1 + 9{,}24 + 0{,}2 + 7{,}92 + 11{,}42 + +10{,}12}{46} = \frac{91{,}76}{46} = 2$$

DESVIACIÓN TÍPICA O ESTÁNDAR

Indica qué tan dispersos están los valores del conjunto de datos respecto a su valor promedio expresada por s, es una medida de cuánto se desvían los valores de datos de la media. Un valor alto de desviación estándar indica que los datos están muy dispersos, mientras que un valor bajo indica que los datos están agrupados cerca de la media (Lozano & Domich, 2023).

Algunos ejemplos de su uso son:

Comparar la variabilidad de diferentes conjuntos de datos, grupos o categorías.

Se utiliza para determinar la confiabilidad de un instrumento de medición o la precisión de un experimento.

Analizar el riesgo en inversiones en finanzas

Controlar la calidad en procesos industriales, se utiliza para monitorizar la variabilidad en la producción y detectar posibles problemas de calidad.

La desviación estándar es la raíz cuadrada positiva de la varianza.

s desviación estándar muestral

σ desviación estándar poblacional

Fórmula:

Para datos No agrupados	
Para una población $$\sigma = \sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$$	Para una muestra $$s = \sqrt{\frac{\sum f(x-\bar{x})^2}{n-1}}$$
Para datos Agrupados	
Para una población $$\sigma = \sqrt{\frac{\sum f(X-\bar{x})^2}{N}}$$	Para una muestra $$s = \sqrt{\frac{\sum f(X-\bar{x})^2}{n-1}}$$

Coeficiente de Variación (Cv)

Es la relación entre la desviación típica y su media. El coeficiente de variación se suele expresar en porcentaje.

$$Cv = \frac{\sigma}{\bar{x}} * 100\%$$

Donde:

σ = *Desviación típica*

$\bar{x}$ = *media aritmética*

Ejemplo:

Los datos descritos en la siguiente tabla corresponden a una muestra de la edad de los pacientes que asisten a un consultorio médico durante un mes. Determine la desviación media, la varianza , la desviación típica y el coeficiente de variación.

Edad años	xi (marca de clase)	fi	xi.fi	$\lvert xi-\bar{x}\rvert$	$\lvert xi-\bar{x}\rvert.fi$	$(xi-\bar{x})^2$	$(xi-\bar{x})^2.fi$
[20-30)	25	20	500	24,65	493	607,62	12152,4
[30-40)	35	35	1225	14,65	512,75	214,62	7511,7
[40-50)	45	50	2250	4,65	232,5	21,62	1081
[50-60)	55	49	2690	5,35	262,15	28,62	1402,38
[60-70)	65	25	1625	15,35	383,75	235,62	5890,5
[70-80)	75	15	1125	25,35	380,25	642,62	9639,3
[80-90]	85	6	510	35,35	212,1	1249,62	7497,72
Total		**200**	**9930**		**2476,5**		**45175,3**

Determinamos el promedio

$$\bar{x} = \frac{\sum xi.fi}{N} = \frac{9930}{200} = 49{,}65$$

Determinamos la desviación media

$$Dm = \frac{\sum \lvert xi-\bar{x}\rvert.fi}{N} = \frac{2476{,}5}{200} = 12{,}38$$

Determinamos la varianza

$$s^2 = \frac{\sum (xi-\bar{x})^2.fi}{n-1} = \frac{45175{,}3}{200-1} = 227{,}01$$

Determinamos la desviación típica

$$s = \sqrt{\frac{\sum (xi-\bar{x})^2.fi}{n-1}} = \sqrt{227{,}01} = 15{,}06$$

Determinamos el coeficiente de variación

$$Cv = \frac{s}{\bar{x}} = \frac{15{,}06}{49{,}65} = 0{,}303 = 30{,}3\%$$

UNIDAD 3

PROBABILIDAD

La probabilidad es una herramienta esencial en muchos campos, como la estadística, la ingeniería, las finanzas, la ciencia de datos y la inteligencia artificial, se utiliza para analizar datos, predecir resultados, tomar decisiones bajo riesgo y evaluar la confiabilidad de los modelos, como disciplina matemática permite cuantificar y comprender la incertidumbre inherente a eventos aleatorios (Vera, 2017).

En estas secciones se estudiarán la regla de Probabilidad a priori a través del cálculo de probabilidades por la regla de Laplace, posterior la probabilidad condicional, reglas de suma y multiplicación y finalmente la Distribución Binomial y de Poisson.

¿QUÉ ES LA PROBABILIDAD?

El concepto de la probabilidad no es ajeno al campo de la ciencia: cuando los resultados de los experimentos no pueden predecirse con exactitud, es importante disponer al menos de una medida del grado de certidumbre con que puede ocurrir cada uno de sus posibles resultados. Esa medida es precisamente lo que llamamos probabilidad (Escudero-Tena et al., 2024).

Se puede mencionar que la probabilidad estudia y determina numéricamente la posibilidad de que ocurra un evento. Se expresa en un rango de 0 (imposible) a 1 (seguro).

Pierre Simon Laplace (1749-1827) define a la probabilidad como una rama de las matemáticas que estudia los acontecimientos inciertos o que dependen del azar.

Para Batanero (2005) es Laplace quien en 1814 introduce el cálculo de la probabilidad en términos de razón o fracción. Es lo que hoy en día se conoce comúnmente como definición (clásica) de probabilidad. El mismo contempla una totalidad de eventos posibles y no posibles. A partir de allí, se busca determinar la probabilidad de ocurrencia de algunos de los eventos, considerando la totalidad de eventos.

CONCEPTOS CLAVES

Evento: Es un suceso que tiene resultados o consecuencias de un procedimiento.

Experimento aleatorio: es aquel evento cuyo resultado es incierto, no se puede determinar con certeza o está sujeto al azar. Por otro lado, un experimento determinista es aquel en el que ya se puede predecir su resultado con anticipación, sin realizarlo.

Evento simple: es cada posible resultado individual de un experimento.

Espacio muestral: asociado a un experimento aleatorio es el conjunto de todos los posibles resultados de un experimento. Representaremos habitualmente el espacio muestral por E.

Ejemplo:

Al lanzar un dado, el conjunto de posibles resultados elementales del experimento es E = {1,2,3,4,5,6}

Suceso seguro: Es aquel que podremos predecir que con seguridad ocurrirá al realizar el experimento aleatorio. Contendrá pues todos los sucesos elementales, por lo que es el propio espacio muestral E.

Ejemplo:

Al lanzar un dado al azar, el Suceso Seguro es "Obtener un número del 1 a 6", es seguro que caerá en cualquiera de esos números

Suceso imposible: Es aquel que podremos predecir que con seguridad no ocurrirá. Como conjunto no contendrá a ningún suceso elemental, por lo que se trata del conjunto vacío, el cual representaremos por ∅.

Ejemplo:

Al lanzar un dado al azar, el Suceso "Obtener un número mayor que 6" es un suceso imposible.

CÁLCULO DE PROBABILIDADES

Determinar la probabilidad de que suceda un evento es encontrar mediante un proceso matemático que tan probable o posible es que suceda dicho evento. A continuación, se presenta algunas reglas para calcular la probabilidad dependiendo de las condiciones y estructura del problema.

REGLA DE LAPLACE

Se define la probabilidad frecuentista o empírica o a priori de un suceso A, representada por P (A) como el valor obtenido para la frecuencia relativa con que se observa A, en un número grande de repeticiones del experimento. Respecto a esta definición se tiene la Regla de Laplace que se aplica cuando todos los resultados en el espacio muestral son igualmente probables.

$$P(A) = \frac{\textit{Número de resultados favorables}}{\textit{Número total de resultados}}$$

Proceso para determinar la probabilidad de un suceso o evento:

1. Definir el evento de interés: identificar claramente el evento cuya probabilidad se quiere determinar. Este evento puede ser cualquier suceso que tenga un resultado incierto, como obtener una cara específica al lanzar una moneda, sacar un número determinado al lanzar un dado, o que un equipo gane un partido de fútbol.

2. Identificar el espacio muestral: es el conjunto de todos los resultados posibles del experimento o situación aleatoria. Es importante **considerar todos los resultados posibles,** sin omitir ninguno. Por ejemplo, al lanzar una moneda, el espacio muestral sería {cara, cruz}.

3. Determinar los casos favorables: son los resultados a favor del espacio muestral que cumplen con las condiciones del evento de interés. Por ejemplo, si queremos saber la probabilidad de obtener cara al lanzar una moneda, el caso favorable sería "cara".

4. Calcular la probabilidad: Aplicar la regla de Laplace.

5. Interpretar el resultado: el valor obtenido para la probabilidad representa la posibilidad de que ocurra el evento de interés. Una probabilidad de 0 indica que el evento es imposible, mientras que una probabilidad de 1 indica que el evento es seguro. Valores entre 0 y 1 representan la frecuencia relativa del evento.

Ejemplo 1:

Se lanza un dado justo de seis caras. ¿Cuál es la probabilidad de obtener un 3?



Solución:

Identificar:

Número de resultados favorables **solo hay un 3**, entonces es 1caso favorable

Número toral de resultados **son 6** porque 6 caras tiene un dado

Aplicamos la regla de Laplace:

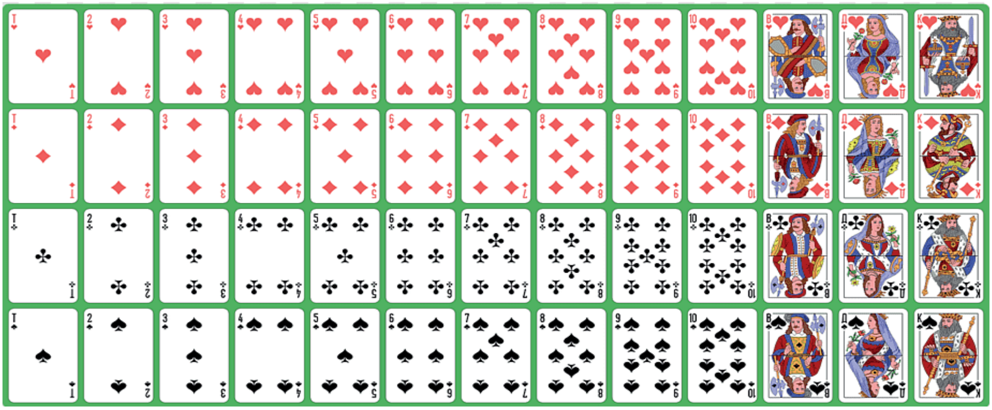
$$P\ (A) = \frac{\textit{Número de resultados favorables}}{\textit{Número total de resultados}}$$

$$P\ (E) = \frac{1}{6} = 0{,}166 = 16{,}6\%$$

Interpretación: La probabilidad de que al lanzar un dado se obtenga un **3** es del 16,6%

Ejemplo 2:

Tenemos una baraja de cartas común con 52 cartas. Se desea calcular la probabilidad de obtener una carta de corazones al azar.



Evento de interés: Obtener una carta de corazones.

Espacio muestral: Todas las cartas de la baraja, es decir, 52 cartas. (total de casos)

Casos favorables: 13 cartas de corazones.

Cálculo de la probabilidad:

$$P\ (A) = \frac{\textit{Número de resultados favorables}}{\textit{Número total de resultados}}$$

$$P\ (\textit{carta de corazón}) = \frac{13}{52} = \frac{1}{4} = 0{,}25 = 25\%$$

Interpretación: La probabilidad de obtener una carta de corazones al azar es de 1/4, o lo que es lo mismo, hay un 25% de posibilidades de que ocurra este evento.

OPERACIONES CON SUCESOS

Un **evento compuesto** es cualquier evento que combina dos o más eventos simples.

UNIÓN DE SUCESOS Y REGLA DE LA ADICIÓN

La regla de la adición o regla de la suma establece que la probabilidad de ocurrencia de cualquier evento en particular es igual a la suma de las probabilidades individuales, si es que los eventos son mutuamente excluyentes, es decir, que los dos no pueden ocurrir al mismo tiempo (Vásquez & Alsina, 2023).

La palabra **o** está asociada con la adición de probabilidades.

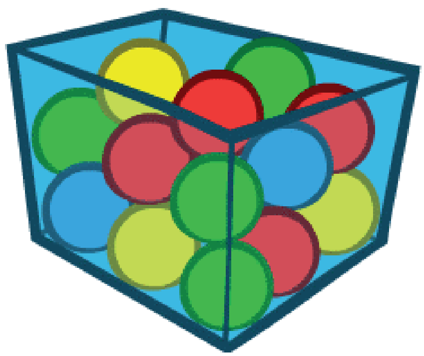
Expresión matemática

P(A o B) = P(A) U P(B) = P(A) + P(B) si A y B son mutuamente excluyente, donde A y B no pueden ocurrir al mismo tiempo.

P(A o B) = P(A) + P(B) – P(Ay B) si A y B son no excluyentes, donde P(Ay B) es la probabilidad de que A y B ocurran al mismo tiempo.

Ejemplo:

En una urna se tienen 4 bolas rojas, 3 amarillas, 4 verdes y 2 celestes. Encuentre la probabilidad de que al extraer una bola esta sea roja o azul.



Solución:

Determinamos el total de casos: sumamos las bolas colores: 4r+3a+4v+2a = **13** bolas en total

Casos favorables (bola roja) = 4

Casos favorables (bola azul) = 2

La palabra clave es **o** por lo tanto se debe aplicar la regla de la suma de probabilidades

Se determina que son mutuamente excluyentes ya no hay la posibilidad de que una bola al mismo tiempo sea roja y azul

Aplicamos la fórmula:

P (roja o azul) = P(R) + P (A)

$$P\,(roja\ o\ azul) = P\,(R) + P\,(A)$$

$$P\,(roja\ o\ azul) = \frac{4}{13} + \frac{2}{13}$$

$$P\,(roja\ o\ azul) = \frac{6}{13} = 0{,}4615 = 46{,}15\%$$

La probabilidad de que el al sacar una bola esta sea de color rojo o azul es del 33,33%.

Ejemplo 2:

¿Cuál es la probabilidad de que al lanzar un dado se obtenga un número par o múltiplo de 3?



Solución

Casos posibles E = {1, 2, 3, 4, 5, 6} 6 casos posibles

Casos favorables P(número par) = 2,4,6 llamaremos P(A) 3 casos favorables

Casos favorables P(múltiplo de 3) = 3,6 llamaremos P(B) 2 casos favorables

Evento (A y B) Identificamos que no son excluyes, ya que el **6** es a la vez par **y** múltiplo de 3

Aplicamos la fórmula:

$$P\,(A\ o\ B) = \underline{P\,(A)} + \underline{P\,(B)} - \underline{P\,(A y\ B)}$$

$$P\,(A\ o\ B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6}$$

$$P\,(A\ o\ B) = \frac{4}{6} = \frac{2}{3} = 0,6666 = 66,66\%$$

La probabilidad de que al lanzar un dado este se un número par o múltiplo de tres es del 66,66%

INTERSECCIÓN DE SUCESOS Y REGLA DE LA MULTIPLICACIÓN

Generalmente se emplea esta regla para eventos sucesivos, como por ejemplo lanzar una moneda dos veces o más, extraer dos bolas una tras otra.

La regla de la multiplicación establece que la probabilidad de ocurrencia de dos o más eventos estadísticamente independientes, que es igual al producto de sus probabilidades individuales.

La palabra **y** está asociada a la multiplicación de probabilidades.

Notación:

***P* (*A y B*)** es la probabilidad de que (el evento A ocurra en un primer ensayo y el evento B ocurra en un segundo ensayo)

***P* (*B*|*A*)** representa la probabilidad de que ocurra el evento B después de suponer que ya haya ocurrido el evento A.

$\boldsymbol{P(A \text{ y } B) = P(A\,B) = P(A) \,.\, P(B)}$ *si A y B son independientes.*

$\boldsymbol{P(A \text{ y } B) = P(A\,B) = P(A).\, P(B|A)}$ *si A y B son dependientes.*

Precaución: La notación P(AyB) tiene dos significados, dependiendo de su contexto. Para la regla de la multiplicación, P(AyB) expresa que el evento A ocurre en un ensayo y el evento B ocurre en otro; para la regla de la suma, utilizamos P(AyB) para indicar que los eventos A y B ocurren ambos en el mismo ensayo.

Eventos independientes: si la ocurrencia de uno no afecta la probabilidad del otro.

Eventos dependientes: si la ocurrencia de uno afecta al otro.

Ejemplo:

Evento dependiente

Una caja contiene 5 canicas azules y 3 rojas. Si se extraen **dos canicas al azar sin reposición**, ¿cuál es la probabilidad de que las dos sean rojas?

Solución:

Las canicas serán extraídas de la misma caja una tras otra y no serán devueltas a la caja (no hay reposición), entonces, se trata de eventos dependientes.

Total de casos posibles **8** en un principio

Evento A: obtener una canica roja en la primera extracción **3 de 8**

$$P(A) = \frac{3}{8}$$

Evento B: obtener una canica roja en la segunda extracción **2 de 7**

$$P(A) = \frac{2}{7}$$

Aplicamos la fórmula:

$P(A \text{ y } B) = P(A\,B) = P(A) \,.\, P(B|A)$

$$P(A \text{ y } B) = \frac{3}{8} \cdot \frac{2}{7}$$

$$P(A\ y\ B) = \frac{6}{56} = \frac{3}{28} = 0{,}1071 = 10{,}71\%$$

Evento independiente

En una Unidad Educativa la probabilidad de que un estudiante seleccionado al azar hable inglés es de 0,20; mientras que la probabilidad de que un alumno juegue fútbol es de 0,80.

El hecho de que un alumno hable inglés, no afecta en nada que juegue fútbol; por lo tanto, se trata de eventos independientes.

Evento A: que el alumno hable inglés $P(A) = 0{,}20$

Evento B: que el alumno juegue fútbol $P(B) = 0{,}80$

Solución:

$P(A\ y\ B) = P(A\ B) = P(A)\ .\ P(B)$

$P(A\ y\ B) = (0{,}20)\ .\ (0{,}80)$

$P(A\ y\ B) = 0{,}16 = 16\%$

PROBABILIDAD CONDICIONAL

La probabilidad condicional se refiere a la probabilidad de que ocurra un evento A dado que otro evento B ya ha ocurrido, se denota como (P I A) y se lee como "la probabilidad de A dado B"

Sean A y B eventos tales que P(B) > 0, la probabilidad del evento A condicional a la ocurrencia del evento B es:

$$P(A\,|\,B) = \frac{P(A \cap B)}{P(B)}$$

Ejemplo:

Al 25% de tus amigos le gusta la fresa y el chocolate, mientras que al 60% le gusta el chocolate. ¿Cuál es la probabilidad de que a un amigo **que le gusta el chocolate, le guste la fresa**?

Solución:

Se tienen 2 eventos: que a un amigo le guste la fresa, y que a un amigo le guste el chocolate.

Evento A: que a un amigo le gusten los fresa. P(A) = ?

Evento B: que a un amigo le guste el chocolate. $P(B) = 60\% = 0,6$

Evento A y B: que a un amigo le guste la fresa y el chocolate. $P(A \cap B) = 25\% = 0,25$

Calculamos la probabilidad de que a un amigo le guste la fresa, dado que le gusta el chocolate.

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

$$P(A \mid B) = \frac{0,25}{0,6}$$

$$P(A \mid B) = \frac{5}{12}$$

$$P(A \mid B) = 0,4167$$

$$P(A \mid B) = 41,67\%$$

La probabilidad de que a un amigo le guste la fresa dado que le gusta el chocolate es del 41,67 %.

UNIDAD 4

DISTRIBUCIÓN DE PROBABILIDAD DE VARIABLE DISCRETA

DISTRIBUCIÓN DE PROBABILIDAD

DISTRIBUCIÓN BINOMIAL

La distribución de probabilidad binomial es una distribución discreta que tiene muchísimas aplicaciones, se asocia con un experimento de múltiples pasos que se llama experimento binomial, la palabra binomial viene de otra palabra que significa “dos nombres” y esto nos hará recordar que en cada ensayo que veremos en este tema siempre habrá dos resultados: éxito y fracaso, si es que respondes una pregunta de alternativas al azar, la respuesta es correcta o incorrecta, si es que realizas un control de calidad a un producto, este será defectuoso o no defectuoso, si es que lanzas una moneda, sale cara o sale cruz (Ramirez & Fiestas, 2023).

Una distribución binomial cumple con cuatro requisitos:

- El procedimiento tiene un número fijo de ensayos. (Un ensayo es una sola observación).
- Los ensayos son independientes, lo que significa que el resultado de cualquier ensayo individual no afecta las probabilidades en los otros ensayos.
- Cada ensayo debe tener todos los resultados clasificados en exactamente dos categorías, comúnmente llamadas éxito y fracaso.
- La probabilidad de un éxito se conserva igual en todos los ensayos.

Análisis en un experimento binomial

Un experimento binomial es un experimento que cumple las siguientes condiciones:

- El experimento consta de una secuencia de n ensayos idénticos.
- En cada ensayo hay dos resultados posibles. A uno de ellos se le llama éxito y al otro, fracaso.
- La probabilidad de éxito es constante de un ensayo a otro, nunca cambia y se denota por p. Por ello, la probabilidad de fracaso será $q = (1 - p)$. Esto se debe a que la probabilidad de éxito más la probabilidad de fracaso suman 1.
- Los ensayos son independientes, de modo que el resultado de cualquiera de ellos no influye en el resultado de cualquier otro ensayo.

Antes de resolver un ejercicio aplicando la fórmula de probabilidad binomial, se debe verificar siempre que se cumplen estas cuatro condiciones, pues esta fórmula solo funciona para experimentos binomiales.

Función de probabilidad binomial (fórmula)

$$P(x) = \frac{n!}{x!\,(n - x)!} \cdot p^{x} \cdot q^{n-x}$$

Para $x = 0, 1, 2, ..., n$

EL FACTORIAL DE UN NÚMERO N!

El factorial de un número, representado por el símbolo “!” después del número, se define como el producto de todos los números enteros positivos menores o iguales a ese número. En otras palabras, multiplicamos todos los números enteros positivos desde 1 hasta el número en cuestión.

Por ejemplo:

El factorial de 5 (escrito como 5!) se calcula así: 5! = 5 x 4 x 3 x 2 x 1 = 120

El factorial de 3 (escrito como 3!) se calcula así: 3! = 3 x 2 x 1 = 6

El factorial de 1 (escrito como 1!) se define como 1, ya que cualquier número multiplicado por 1 sigue siendo ese mismo número.

Ejemplo1:

Dado que hay una probabilidad de 0.85 de que un adulto seleccionado al azar sepa lo que es Twitter, use la fórmula de probabilidad binomial para encontrar la probabilidad de que, cuando cinco adultos se seleccionan al azar, exactamente tres sepan qué es Twitter. Es decir, aplique la fórmula para encontrar P(3) dado que $n = 5$, $x = 3$, $p = 0.85$ *y* $q = 0.15$

Solución:

$$P(x) = \frac{n!}{x!\,(n - x)!} \cdot p^{x} \cdot q^{n-x}$$

$$P(x) = \frac{5!}{3!\,(5 - 3)!} \cdot (0{,}85)^{3} \cdot (0{,}15)^{5-3}$$

$$P(x) = \frac{5!}{3!\,(2)!} \cdot (0{,}85)^{3} \cdot (0{,}15)^{2}$$

$P(x) = 0{,}138$

$P(x) = 13{,}8\%$

La probabilidad de obtener exactamente tres adultos que conozcan twitter entre los cinco adultos seleccionados al azar es del 13,8%

Ejemplo 2:

La probabilidad de que a un cliente nuevo le guste la rosahamburguesa de Pancho´s food es de 0,8. Si llegan 5 clientes nuevos a la cafetería, ¿cuál es la probabilidad de que solo a 3 de ellos les guste la rosahamburguesa?

Solución:

Antes de aplicar la fórmula, se debe verificar que se trate de un experimento binomial. Para ello, tiene que cumplir con las 4 condiciones que mencionamos arriba. Efectivamente, se trata de un experimento binomial.

$n = 5$

$x = 3$

$p = 0{,}8$

$q = 1 - p$

$q = 1 - 0{,}8$

$q = 0{,}2$

Aplicar la fórmula binomial:

$$\boldsymbol{P(x)= \frac{n!}{x!\,(n-x)!} \,.\, p^{x} \,.\, q^{n-x}}$$

Reemplazamos estos valores en la fórmula:

$$P(x)= 3 \quad \frac{5!}{3!\,(5-3)!} \,.\, (0{,}8)^{3} \,.\, (0{,}2)^{5-3}$$

$$P(x)= 3 \quad \frac{5.4.3.2.1}{3.2.1\,(2.1)} \,.\, (0{,}8)^{3} \,.\, (0{,}2)^{2}$$

$$P(x)= 3 \quad \frac{5.2.1}{1} \,.\, (0{,}8)^{3} \,.\, (0{,}2)^{2}$$

$$P(x)= 3 \quad 10 \,.\, (0{,}8)^{3} \,.\, (0{,}2)^{2}$$

$$P(x)= 3 \quad 0{,}2048$$

$$P(x)= 3 \quad 20{,}48\%$$

La probabilidad de que de los cincos clientes a tres de ellos les guste la rosahamburguesa es del 20,48%.

DISTRIBUCIÓN DE POISSON

La distribución de Poisson, introducida por el matemático francés Siméon Denis Poisson en 1837, es una distribución de probabilidad discreta que modela la ocurrencia de eventos aleatorios en un intervalo fijo, como el tiempo, la distancia o el área, a lo largo de su historia, ha encontrado un amplio espectro de aplicaciones en diversos campos, desde la física y la ingeniería hasta la biología y la economía (Otero-Potosi, 2023).

Por ejemplo, son variables de Poisson: el número de llamadas que recibe una central telefónica en el período de 1 minuto, el número de bacterias en un volumen de 1 litro de agua o el número de fallas en la superficie de una pieza de cerámica rectangular.

Aplicaciones:

La distribución de Poisson encuentra aplicaciones en una amplia gama de escenarios donde se observan eventos aleatorios e independientes en un intervalo específico. Algunos ejemplos notables incluyen:

Telecomunicaciones: Predecir el número de llamadas telefónicas que recibe una central en un minuto.

Biología: Modelar la cantidad de bacterias presentes en una muestra de agua.

Manufactura: Estimar el número de defectos en una pieza de metal durante el proceso de producción.

Finanzas: Analizar la frecuencia de transacciones en un mercado bursátil.

Tráfico: Predecir el número de vehículos que pasan por una intersección en una hora determinada.

Características:

La distribución de Poisson se caracteriza por las siguientes propiedades:

Eventos aleatorios e independientes: La ocurrencia de un evento no afecta la probabilidad de que ocurra otro.

Intervalo fijo: El número de eventos se observa en un intervalo de tiempo, distancia, área o volumen específico.

Tasa media constante: La tasa promedio de eventos por unidad de intervalo es constante.

Distribución discreta: Los valores que puede tomar la variable aleatoria (número de eventos) son enteros no negativos.

Ecuación de la Función de Poisson

$$f(x) = P(X - x) = \frac{e^{-u}.\mu^{x}}{x!} \quad ; x = 0; 1; 2; 3; 4,5; \ldots$$

Donde:

$f(x) = P(X - x)$ es la probabilidad de x ocurrencias en un intervalo.

u es el valor esperado o media de X

e es la base del logaritmo natural cuyo valor es 2,71828…

Nota

En una distribución de Poisson:

La media de X es igual a *u*

La varianza de X es igual a *u*

Ejemplo:

Una empresa recibe un promedio de 10 llamadas telefónicas en 60 minutos. Se desea calcular la probabilidad de que reciba exactamente 3 llamadas en un intervalo de 30 minutos.

Análisis y solución:

El promedio de eventos por unidad de intervalo es constante, es decir 10 llamadas por hora; la tasa para un intervalo de 30 minutos sería 3 llamadas por 30 minutos es decir 0,5 horas.

Determinamos:

Primero la variable aleatoria:

X= número de llamadas por hora

Segundo calculamos la probabilidad de que se reciban 3 llamadas en 0,5 horas, es decir $f(3)$

En el problema se muestra que en promedio ingresan 10 llamadas en 1 hora, entonces $u = 5$

Aplicamos la fórmula de la distribución de Poisson:

$$f(x) = P(X - x) = \frac{e^{-u}.\mu^{x}}{x!}$$

$$f(3) = P(X = 3) \quad \frac{e^{-5}.5^{3}}{3!} = 0,1403 = 14,03\%$$

Por lo tanto, la probabilidad que la empresa espera que en media se reciba exactamente 3 llamadas es del 14,03%.

DISTRIBUCIÓN HIPERGEOMÉTRICA

La distribución hipergeométrica fue introducida por Jacob Bernoulli, un matemático suizo, en su obra “Ars Conjectandi” publicada en 1713. Este trabajo es uno de los textos fundamentales en el desarrollo de la teoría de la probabilidad.

En teoría de la probabilidad y estadística, la distribución hipergeométrica es una distribución de probabilidad discreta relacionada con muestreos aleatorios y sin reemplazo.

Fórmula:

La probabilidad de obtener exactamente k éxitos en n extracciones sin reemplazo de una población de tamaño N que contiene K éxitos se da por:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Donde:

$\binom{K}{k}$ es el número de combinaciones de K elementos tomados de k en k.

$\binom{N-K}{n-k}$ es el número de combinaciones de N–K elementos tomados de n–k en n–k.

$\binom{N}{n}$ es el número de combinaciones de N elementos tomados de n en n.

APLICACIONES DE LA DISTRIBUCIÓN HIPERGEOMÉTRICA

La distribución hipergeométrica se aplica en situaciones donde se hacen extracciones sin reemplazo. Algunos ejemplos incluyen:

- **Calidad y Control**: En control de calidad, se puede usar para determinar la probabilidad de obtener un cierto número de artículos defectuosos en una muestra de productos de un lote.
- **Juegos de Azar**: En juegos de cartas, se puede usar para calcular la probabilidad de obtener una mano específica en un juego de póker.
- **Biología**: En estudios de genética, se puede usar para calcular la probabilidad de seleccionar un cierto número de individuos con una característica genética particular de una población.
- **Auditorías**: En auditorías financieras, se puede usar para determinar la probabilidad de encontrar un número específico de discrepancias en una muestra de transacciones.

La distribución hipergeométrica es particularmente útil en situaciones donde la población es finita y el tamaño de la muestra es una fracción significativa del tamaño de la población.

Ejemplo:

Control de calidad

Una fábrica produce un lote de 1000 piezas, de las cuales 50 son defectuosas. Se selecciona una muestra de 10 piezas sin reemplazo. Queremos encontrar la probabilidad de que exactamente 2 piezas sean defectuosas en la muestra.

Datos del problema:

- Tamaño de la población (N) = 1000
- Número de éxitos en la población (K) = 50
- Tamaño de la muestra (n) = 10
- Número de éxitos en la muestra (k) = 2

Solución:

Usamos la fórmula de la distribución hipergeométrica:

$$P(X=k)=\frac{\binom{K}{k}\binom{N \quad -K}{n \quad -k}}{\binom{N}{n}}$$

Primero, calculamos los coeficientes binomiales

$$\binom{K}{k}=\binom{100}{2}=\frac{100!}{2!\,(100-2)!}=\frac{100*99}{2*1}=4950$$

$$\binom{N-K}{n-k}=\binom{900}{8}=\frac{900!}{8!\,(900-2)!}\approx 2.160*10^{16}$$

$$\binom{N}{n}=\binom{1000}{10}=\frac{1000!}{10!\,(1000-10)!}\approx 2.631*10^{20}$$

Finalmente, sustituimos estos valores en la fórmula:

$$P(X=2)=\frac{4950*2.160*10^{16}}{2.631*10^{20}}\approx 0{,}0407$$

Por lo tanto, la probabilidad de que exactamente 2 bombillas sean defectuosas en la muestra es aproximadamente 0.0407, o 4.07%.

REFERENCIAS

Aguilar Fernández, E., Andrey Zamora Araya, J., & Zamora, A. Y. (2024). Las tarjetas como recurso didáctico para la enseñanza de la estadística: La experiencia en un curso de didáctica específica. *Innovaciones Educativas*, *26*(40), 135–146. https://doi.org/10.22458/ie.v26i40.4674

Alban, G. P. G., Arguello, A. E. V., & Molina, N. E. C. (2020). Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción). *RECIMUNDO*, *4*(3), 163–173. https://doi.org/10.26820/RECIMUNDO/4.(3).JULIO.2020.163-173

Aquije, A. J. Z., Cardenas, A. J. M., & Salcedo, K. J. P. (2023). Estadística y medicina. Un enfoque para principiantes. *Religacion Press*. https://doi.org/10.46652/RELIGACIONPRESS.101

Casas Anguita, J., Repullo Labrador Donado Campos, J. J., & Casas Anguita, J. (2003). La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos (I). *Atención Primaria*, *31*, 527–538. https://doi.org/10.1016/S0212-6567(03)70728-8

Castañeda, M. B., Cabrera, A., Navarro, Y., & De Vries, W. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS* (1st ed., Vol. 1). EDIPUCRS. https://www.researchgate.net/profile/Alberto-Cabrera/publication/261704346_Procesamiento_de_datos_y_analisis_estadisticos_utilizando_SPSS_Un_libro_practico_para_investigadores_y_administradores_educativos/links/00b4953510e4a0dd01000000/Procesamiento-de-datos-y-analisis-estadisticos-utilizando-SPSS-Un-libro-practico-para-investigadores-y-administradores-educativos.pdf

Chance, B. L. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910677

Cujba, A., & Pifarré, M. (2024). Validación exploratoria de un cuestionario de actitudes hacia la estadística con tecnología. *Campus Virtuales*, *13*(1), 47. https://doi.org/10.54988/CV.2024.1.1266

del Pino, G., & Estrella, S. (2012). Educación estadística: relaciones con la matemática. *Pensamiento Educativo: Revista de Investigación Educacional Latinoamericana*, *49*(1), 53–64. https://doi.org/10.7764/PEL.49.1.2012.5

Dexeus, C. R. (2001). La Estadística de movimientos turísticos en fronteras (FRONTUR): oportunidad de negocio. *Estudios Turísticos*, *148*(148), 69–90. https://doi.org/10.61520/ET.1482001.866

Escudero-Tena, M., Ojeda-Casares, S., Moya, L.-Á., & Enrique-Regueira, I. (2024). La malla estadística como unidad de análisis espacial. Razón de mortalidad, población y vivienda. *Revista EURE - Revista de Estudios Urbano Regionales*, *50*(150). https://doi.org/10.7764/EURE.50.150.11

Fernández Altamirano, A. E. F., & Vela Meléndez, L. (2021). *Los paradigmas y las metodologías usadas en el proceso de investigación: una breve revisión*. http://rua.ua.es/dspace/handle/10045/119978

Franco, M. F. de, & Solórzano, J. L. V. (2020). Paradigmas, enfoques y métodos de investigación: análisis teórico. *MUNDO RECURSIVO*, *3*(1), 1–24. https://www.atlantic.edu.ec/ojs/index.php/mundor/article/view/38

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459. https://doi.org/10.1038/NATURE14541

Giuliana, B., & Silvana, M. M. (2023). Estadística en la formación y práctica médica. *Revista de La Facultad de Ciencias Médicas. Universidad Nacional de Rosario.*, *3*, 21–28. https://doi.org/10.35305/FCM.V3I.111

Guevara Alban, G., Verdesoto Arguello, A., & Castro Molina, N. (2020). Metodologías de investigación educativa (descriptivas, experimentales, participativas, y de investigación-acción). *RECIMUNDO: Revista Científica de La Investigación y El Conocimiento, ISSN-e 2588-073X, Vol. 4, Nº. 3, 2020, Págs. 163-173*, *4*(3), 163–173. https://doi.org/10.26820/recimundo/4.(3).julio.2020.163-173

Guttman, L. (1979). Malos usos en Estadística. *Revista Española de Investigaciones Sociológicas*, *6*, 101–127. https://doi.org/10.5477/CIS/REIS.6.101

Hernández Dávila, E. S., Gallegos Londoño, C. M., & García Mora, F. A. (2023). Estadística descriptiva para el mantenimiento industrial con Python. *Estadística Descriptiva Para El Mantenimiento Industrial Con Python*. https://doi.org/10.33996/CIDE.ECUADOR.EP2636553

Karen, A., Ortega, Z., & Zaldivar-Ortega, A. K. (2024). Áreas de aplicación de la Estadística. *Logos Boletín Científico de La Escuela Preparatoria No. 2*, *11*(21), 29–31. https://doi.org/10.29057/PREPA2.V11I21.11990

Larios Ramirez, O. S., & Fiestas Elías, F. J. (2023). Estadística descriptiva con Excel. *Repositorio Institucional - UCV*. https://repositorio.ucv.edu.pe/handle/20.500.12692/109914

López, C., & Gómez, P. (2023). REVISIÓN CURRICULAR DE LOS TEMAS DE ESTADÍSTICA EN EDUCACIÓN PRIMARIA. *Revista Latinoamericana de Investigación En Matemática Educativa*, *26*(1), 81–100. https://doi.org/10.12802/RELIME.23.2613

Lozano, K. L. G., & Domich, M. A. A. (2023). Relación entre la estadística cívica y la construcción de ciudadanía. *Ciencia Latina Revista Científica Multidisciplinar*, *7*(2), 2694–2709. https://doi.org/10.37811/CL_RCM.V7I2.5518

Millán, C. A. (2023). LA ESTADÍSTICA COMO CONTENIDO FUNDAMENTAL EN EL ÁREA DE MATEMÁTICAS. *DIALÉCTICA*, *2*(22). https://doi.org/10.56219/DIALCTICA.V2I22.2691

Monsalve, L. (2024). ENSEÑANZA DE LA ESTADÍSTICA EN LA FORMACIÓN PROFESIONAL DE INGENIEROS. *LÍNEA IMAGINARIA*, *1*(18). https://doi.org/10.56219/LNEAIMAGINARIA.V1I18.2580

Otero-Potosi, S. A. (2023). Análisis de los procesos de Acreditación de Institutos Técnicos y Tecnológicos en el Ecuador. *Revista Latinoamericana Ogmios*, *3*(8), 1–10. https://doi.org/10.53595/RLO.V3.I8.072

Pérez Juste, R., Galán González, A., & Quintanal Díaz, J. (2012). *Métodos y diseños de investigación en educación*. Universidad Nacional de Educación a Distancia.

Ramírez Noriega, A. D., Tripp Barba, C., & Jiménez Calleros, S. P. (2024). Probabilidad y estadística en la toma de decisiones. *Revista Digital Universitaria*, *25*(2). https://doi.org/10.22201/CUAIEED.16076079E.2024.25.2.5

Rodríguez Cisneros, L. M., Araujo de Rodríguez, I. del C., Navarrete Pilacuán, M. P., & Duque Granados, R. A. (2023). Estadística Aplicada. *Estadística Aplicada*. https://doi.org/10.37811/CLI_W832

Rodríguez-Alveal, F., Díaz-Levicoy Profesor de, D., & Rodríguez-Alveal Danilo Díaz-Levicoy, F. (2024). *ACTITUDES DE FUTUROS PROFESORES DE MATEMÁTICA HACIA LA ESTADÍSTICA Y SU ENSEÑANZA: UNA APROXIMACIÓN A LA ALFABETIZACIÓN ESTADÍSTICA*. *49*(2).

Rumsey, D. J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, *10*(3). https://doi.org/10.1080/10691898.2002.11910678

Vásquez, C., & Alsina, Á. (2023). Creencias del profesorado de educación primaria en torno a la enseñanza de la estadística. *Revista Chilena de Educación Matemática*, *15*(3), 90–101. https://doi.org/10.46219/RECHIEM.V15I3.133

Vera, N. T. J. (2017). Actitud Hacia la Estadística y Estilos de Aprendizaje Hacia la Estadística en los Estudiantes de una Universidad Privada de la Ciudad de Lima. *Revista Científica Pakamuros*, *5*(1). https://doi.org/10.37787/8E5EP246

Vinicio Yanqui Avilés, M., Alexandra Quintana López, X., Telenchana, L., & Avilés, Y. (2024). Determinación de dependencia estadística de la temperatura de trabajo de rodamientos rígidos en procesos industriales mediante ANOVA en RStudio. *ConcienciaDigital*, *7*(1), 82–99. https://doi.org/10.33262/concienciadigital.v7i1.2904



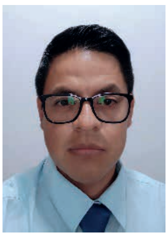
MIRIAN ALEXANDRA VALERIANO MENESES: Licenciada en Ciencias de la Educación especialidad Física y Matemática

Magíster en Tecnología e Innovación Educativa

Magíster en Estrategias para la docencia STEM con mención en ciencias físicas

Doctoranda del programa de Doctorado en Educación e Innovación

https://orcid.org/0009-0006-3089-537X



SANTIAGO ANDRÉS OTERO-POTOSI: Ingeniero en Mantenimiento Automotriz

Magister en Gestión de la Calidad en Educación

Master Universitario en Sistemas Integrados de Gestión

Doctor en Evaluación y Acreditación de Instituciones de Educación Superior

https://orcid.org/0000-0002-3823-9522

INSTITUTO SUPERIOR TECNOLÓGICO
LICEO ADUANERO

© La presente Guía general de estudio de la asignatura de Estadística para Educación Inicial se deriva del proyecto de I+D del instituto Superior Tecnológico Liceo Aduanero denominado "Procesos de Gestión de Calidad en Instituciones de Educación Superior".

GUIA GENERAL DE ESTUDIO DE LA ASIGNATURA DE

ESTADÍSTICA APLICADA

PARA EDUCACIÓN INICIAL

www.atenaeditora.com.br

contato@atenaeditora.com.br

@atenaeditora

www.facebook.com/atenaeditora.com.br

Atena Editora

Año 2024

GUIA GENERAL DE ESTUDIO DE LA ASIGNATURA DE

ESTADÍSTICA APLICADA

PARA EDUCACIÓN INICIAL

www.atenaeditora.com.br

contato@atenaeditora.com.br

@atenaeditora

www.facebook.com/atenaeditora.com.br

Atena Editora

Año 2024